Personal Data Management systems

Master Thesis Computer Science
Radboud University Nijmegen

Author:	Supervisor:
Robin Oostrum	Mireille Hildebrandt
alfroj@gmail.com	hildebrandt@law.eur.nl

17th August 2015

Abstract

In our data-driven society, personal data become more and more valuable every day. Advertising companies aggregate our location histories, phone call data and online clicking behavior into extensive and highly valuable profiles - often without us, the data subjects, noticing. Personal Data Management is needed to gain back control over, and insight in, the processing of such personal data. In this thesis, I look at the legal requirements concerning data processing within existing European data protection and privacy legislation. Discussing and comparing four relevant Personal Data Management systems, I test to what extent the technical specifications of such PDMs are compatible with these legal standards.

Acknowledgements

I would never have finished this thesis without the help of my supervisor, **prof. mr. dr. Mireille Hildebrandt**. She provided me with a relevant research subject and a substantial amount of relevant reading work, and always appeared genuinely interested in my progress. Although she must have wondered whether I would ever finish this thesis, she kept encouraging me to continue. The time she spent in reading and commenting my weekly updates and questions is very much appreciated. Also a big thanks to **Jaap-Henk Hoepman** for taking the time to review my thesis as a second reader and to **Merel Koning** for reading and commenting an earlier concept version.

My research on the Personal Data Management system of Mydex could not have been completed without the help of **William Heath**, who took the time to respond to my e-mails and answered questions that would have remained unanswered without his help. The same goes for **Luk Vervenne** of Synergetics, the quickest responder to e-mails of them all, with whom I also enjoyed an interesting real-life conversation about his own PDM and his view on personal data management in general.

A special thanks to **Imke**, who provided me with a room to work and live in, an endless amount of coffee, food and support, and - during the final stages - helped me create a plan on how to complete this thesis. Finally, I would like to thank my **parents**, who always seemed to sense when it was alright to ask about my progress (and when not), finding the right balance between encouragement and support.

Contents

1 Introduction

1.1 Problem Description

While data are shared everywhere at every minute, it becomes more and more unclear what exactly happens to these often personal and sensitive data. Although we started to realize that sharing photos on social network sites can be ignorant and decrease our chances for future jobs, we often still feel we are in control: deleting such a picture surely is only one click away. Yet slightly more beyond our control are the data these network sites observe from us: our clicking behavior or browsing and location history are often far more interesting for (for example) advertising companies than just a holiday picture on our facebook timelines. Collections of these observed data lead to often huge - highly valuable - data profiles of one person, who often has no clue these data exist, let alone are processed by some third party he or she has never even heard of.

There are laws with which such data controllers and data processors must comply. The EU Data Protection Directive 95/46/EC[1] (from now on referred to as DPD) provides explicit rules on legal grounds for data processing to make sure personal data are protected. Yet the data subjects themselves have little control over data once collected or passed on to third parties. Personal Data Management (PDM) is needed to gain back control over and insight in the processing of personal data. During the past few years, some Personal Data Management systems (PDMs) have been developed to achieve this. It will for example - in the case of data processed on the basis of consent - be interesting to see whether they can really shift the control on personal data back to the user, and how they will protect our personal data from being accessed and/or processed unlawfully. For other legal processing grounds, such as vital interests, they will hopefully improve transparency and ensure that processing is done in accordance with the corresponding legal requirements.

PDMs focus on the management of personal data and their corresponding transparency: access, use and abuse of data, passing data to third parties, the subject's right of access and removal of data. These relate to the DPD, discussing legal grounds and requirements for the processing of personal data. Apart from the legal meaning of personal data, the World Economic Forum ([World Economic Forum, 2012]) distinguishes three types of personal data: *volunteered*, *observed* and *inferred*:

volunteered data refers to data users "explicitly share about themselves". For example: pictures shared on social network sites (Facebook, Instagram, Twitter), personal blogs or billing information during an online purchase.

observed data are data "captured by recording activities of users", such as cell phone location data, search histories or digital cookies.

inferred data are the result of "analysis of personal data" using data mining technologies on (collections of) volunteered and/or observed data. Examples are future consumption prediction techniques or credit scores. In-

[1] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data - `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML`

ferred data need not necessarily be personal data, but might nonetheless have an impact on individuals.

Most business cases are mainly interested in the latter two. This is in contrast with the control the user has: while volunteered data seems manageable and controllable, this sense of ownership shifts towards the data controller when talking about observed and inferred data. This brings us to the goal of a PDM: "that a users personal data are not shared without their consent, or in the case of necessity (contract, a legal obligation, vital interests, public tasks or the legitimate interests of the data controller), on condition that they will not be used for other purposes than those stipulated when access was provided" ([Hildebrandt et al., 2013]).

A common approach for PDMs is to build an entire secure structure to store encrypted personal data. The data can then only be accessed, managed and processed by entities that are (by user and/or law) allowed to. One practical approach is Mydex, a platform that aims to achieve "individual-centric control over personal data and identity" ([Heath et al., 2013]). The main issue with this type of PDM is the lack of control on data when getting out of such a system (that is: after data are decrypted). As [Kuppinger and Kearns, 2013] summarize: "once it's out, it's out of control."

One possible way of ensuring personal data are not being used for other purposes is the use of so-called *sticky policies*: user permissions and policies get permanently bound to the data, ensuring any party processing these data knows the user's wishes and desires. Ideally users are also able to modify their preferences later on and prevent further processing and/or use of previously collected data. Such a metadata-based architecture for sticky policies is described by [Nguyen et al., 2013]. Note that this approach still requires a certain trust framework: in order to make data controllers respect the policies, the legal consequences for violating them must be detected and enforced.

1.1.1 Research Questions

This leads to the following research question:

> **To what extent are the technical specifications of PDMs compatible with the relevant legal standards?**

which can be divided into three subquestions:

1. What EU Data Protection problem(s) can PDMs solve?
2. What EU Data Protection problem(s) can PDMs *not* solve?
3. What new problem(s) do PDMs create?

1.2 Research Method

My research method entirely consists of desk research and literature study, in which I will try to build a bridge between the legal world and the world of computer science and - in particular - digital security. I do this by creating a kind of interface between the legal standards of privacy and data on one hand, and the technical requirements that can (and should) facilitate these

standards on the other hand. Focusing on Personal Data Management systems, I will examine to what extent the technical specifications of such PDMs are compatible with the relevant legal standards.

I will start with a general overview of Personal Data Management systems: what does a typical PDM consist of, how can one be analyzed, who are the stakeholders, et cetera.

Before I can give a proper answer to the research questions in section 1.1.1, I will dive into the legal background of personal data management: what is allowed, what parties are typically involved in data processing, what safeguards are required, what is only allowed under certain circumstances, and so on. I will do so by analyzing the legal framework of the DPD, the proposal for a General Data Protection Regulation (9565/15)[2] (from now on referred to as pGDPR) and possible other relevant directives like the ePrivacy Directive (2002/58/EC). These backgrounds and, more specifically, the legal grounds for data processing, are discussed in section 3. There I will show that there is a lot more to be considered than the well-known "[e]veryone has the right to the protection of personal data concerning him or her", from Article 8 of the EU Charter of Fundamental Rights.

Once I have finished my analysis of the legal framework, I will apply this knowledge on a selection of Personal Data Management systems. This has several advantages: by discussing several PDMs one by one instead of all PDMs in general, I can take a closer look at how each of these PDMs behave and how they rate on legal requirements. I will do this by looking at the different actors and data types involved in each PDM and how they technically ensure these legal requirements are met. Given the broad variety of existing PDMs, my selection will consist of different types: while most PDMs focus on volunteered data, I also want to research the possibilities related to observed and (if possible) inferred data. This has led to a selection of the following PDMs: Mydex, IRMA, Synergetics and openPDS/SafeAnswers. My findings can be found in section 4.

In the end, I will conclude by comparing the pros and cons of the discussed PDMs. This will lead to answers to, as well as analysis of, the research questions: what problems can PDMs solve, what problems cannot be solved and what new problems do they create?

[2]Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) - `http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf`

2 Background

2.1 Analytical Framework of PDMs

A meta-description of Personal Data Management systems is offered by [Bus and Nguyen, 2013]. They divide the main issues related to PDMs into three layers: infrastructure, data management and user interaction. They can be summarized as follows:

Infrastructure of a PDM has two main goals: to assure both the *integrity* and *confidentiality* (defined as *security* by [Bus and Nguyen, 2013]) of the data. This includes supporting all the appropriate encryption techniques, logging and monitoring tools, authentication and identifying protocols etc. Trustworthiness and acceptability of such an infrastructure "may be achieved through market mechanisms (reputation, brands, price), through regulation, certification, control, and enforcement, or through other more direct forms of governance directly supervising (parts of) the infrastructure."

Data management focuses on the ways the PDM ensures a safe and effective control of data, including permission handling mechanisms, communication policies, data auditing possibilities etc. The most common and straight-forward approach to accomplish data management trustworthiness is the signing of a contract between data subject (user) and data controller, making the controller responsible and - more important - accountable in the case of ignoring their given permissions.

User interaction is defined by [Bus and Nguyen, 2013] as "the elements that enable end users to have a meaningful interaction with service providers regarding the permissions and policies associated with the use of their personal data." They suggest that every PDM should offer simple and intuitive tools to control *context-aware data sharing*, which all rely on trustworthy underlying data management and infrastructure layers.

2.1.1 Actors

When discussing the different actors involved in a certain PDM, it is important to distuinguish a legal from an organizational background. The latter will mainly divide the stakeholders into three groups: service providers, (end) users and PDM providers themselves. Article 2 of the DPD on the other hand identifies data controllers, data processors and data subjects:

> (a) 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity; (d) 'controller' shall mean the natural or legal person, public authority, agency or any other body **which** alone or jointly with others **determines the purposes and means of the processing of personal data**; where the purposes and means of

> processing are determined by national or Community laws or regulations, the controller or the specific criteria for his nomination may be designated by national or Community law; (e) 'processor' shall mean a natural or legal person, public authority, agency or any other body **which processes personal data on behalf of the controller**;

These different approaches both differ and overlap. The data subject is the same as the end user of the PDM, as it is their data we are talking about. Both the PDM and the service providers can then be seen as data *controllers*, as they keep and/or process data for specific purposes: providing a service to the end user, even when - in the case of the PDM provider - that service equates to protecting the data against unlawful data processing by other parties.

2.1.2 Implementation

The aforementioned paper by [Bus and Nguyen, 2013] also discusses (chapter 4.5 of [Bus and Nguyen, 2013]) the three key elements of ensuring trust within such trust networks: a one-time technology implementation, signed contracts between the stakeholders and - most important - governance to "ensure proper oversight, auditing, decision-making about members and monitoring procedures, and adapting the rules and conditions to changing circumstances."

Combining multiple trust networks based on a common set of rules would then ultimately lead to a *data ecosystem*, which ideally "would reflect and integrate with offline life in society" and facilitate all three layers described above. Apart from the three layers within PDMs, [Bus and Nguyen, 2013] distinguish ("but not disconnect") five discplines considered within such an ecosystem (or *context-aware personal data management system*). Those disciplines are the *technical*, *economic*, *legal*, *socio-political* and *intergovernmental* perspectives. In this thesis I will mainly focus on the technical and legal aspects.

3 Data Protection & Privacy Law

In this chapter I will summarize the legal framework related to data protection and privacy law. Be aware that data protection and privacy are, although strongly connected, two different fundamental rights. The Charter of Fundamental Rights of the European Union defines those two rights in Article 7 (privacy) and 8 (data protection):

- "Everyone has the right to respect for his or her private and family life, home and communications."[3]
- "Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified."[4]

When talking about personal data management systems, I will focus on data protection law. In this chapter I will elaborate on the summary above by diving into quoted terms such as *consent*, *legitimate basis*, *specified purposes* et cetera. When personal data management systems enable the realization of those data protection rights (and duties), we speak of *data protection by design*. This relates to Article 23 of the pGDPR, which states that data controllers should always implement appropriate technical measures to ensure those data protection rights are honored:

> (1) Having regard to available technology and the cost of implementation and taking account of the nature, scope, context and purposes of the processing as well as the likelihood and severity of the risk for rights and freedoms of individuals posed by the processing, the controllers shall implement technical and organisational measures appropriate to the processing activity being carried out and its objectives, such as data minimisation and pseudonymisation, in such a way that the processing will meet the requirements of this Regulation and protect the rights of data subjects.
>
> (2) The controller shall implement appropriate measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed; this applies to the amount of data collected, the extent of their processing, the period of their storage and their accessibility. Where the purpose of the processing is not intended to provide the public with information, those mechanisms shall ensure that by default personal data are not made accessible without human intervention to an indefinite number of individuals.

This implies that data controllers should use a PDM if such a technology ensures that the requirements of the data regulations are met. Potentially this

[3] Article 7 of the EU Charter of Fundamental Rights - `http://www.eucharter.org/home.php?page_id=14`

[4] Article 8 of the EU Charter of Fundamental Rights - `http://www.eucharter.org/home.php?page_id=15`

could lead to PDMs being implemented on a wide scale, protecting the rights of the data subjects. I will now further discuss the relevant rights and duties in the following paragraphs.

3.1 Legal Grounds

3.1.1 Consent

To process personal data, one of several legal grounds has to be satisfied. Article 7 of the DPD[5] specifies those legal grounds. One of them is *consent* from the data subject. Article 2 of the same directive defines the data subject's consent as "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed." However, the DPD lacks any further definition of *freely given*, *specific*, *informed* and *unambiguously*. A more detailed - 38 page - explanation was written by the Article 29 Working Party in Opinion 15/2011 on consent[6]. From this article, the following definitions can be derived:

freely given means that "there must be no risk of deception, intimidation or significant negative consequences for the data subject if he/she does not consent". It is for example doubtful whether consent is given entirely freely in an employment environment when there is an element of subordination. Recital 34 of the pGDPR adds that "[i]n order to safeguard that consent has been freely-given, consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller and this imbalance makes it unlikely that consent was given freely in all the circumstances of that specific situation. Consent is presumed not to be freely given, if it does not allow separate consent to be given to different data processing operations despite it is appropriate in the individual case, or if the performance of a contract is made dependent on the consent despite this is not necessary for such performance and the data subject cannot reasonably obtain equivalent services from another source without consent." In other words: refusing consent to the processing of unnecessary data is *not* a valid ground for the data controller to abort (further) provision of a service or execution of a contract. This addition is particularly relevant for mobile apps, which are increasingly asking for contacts, call histories and other data not relevant for executing the app itself.

specific means that "blanket consent without determination of the exact purposes" is not valid. Contracts should provide specific consent clauses, instead of inserting consent-related information in the general conditions. The Article 29 Working Party adds that specific consent also means it must be intelligible: "it should refer clearly and precisely to the scope and the consequences of the data processing. It cannot apply to an open-ended set of processing activities. This means in other words that the context in which consent applies is limited."

[5]DPD - `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML`

[6]All cited sources from the Article 29 Working Party are listed in section A

informed is already defined by Article 10 and 11 of the DPD. They state that Member States shall provide that the data controller must inform the data subject with the identity of the controller and the purposes of the processing, as well as any possible further information such as the recipients of the data. Futhermore, this information should be provided in appropriate language, without the use of "overly complicated legal or technical jargon." It should also be provided directly to the data subject, and not just be available somewhere.

unambiguous consent is mentioned in Article 7(a), which "calls for the use of mechanisms to obtain consent that leave no doubt as to the individual's intention to provide consent."

explicit consent, mentioned in Article 8.2(a), is required to process sensitive data. It means "an active response, oral or in writing, whereby the individual expresses his/her wish to have his/her data processed for certain purposes." A pre-ticked box can never lead to explicit consent being given, as it does not involve some positive action from the data subject.

It is important to note that consent can be revoked (by the user) at any time. This means that - when consent has been revoked - the stored personal data must be anonymised and/or deleted. I will take a closer look at purpose limitation in section 3.2.

3.1.2 Contract

Personal data may also be processed if at least one of the other legal grounds - as specified by Article 7 - is applicable. One of them is a binding contract between the data subject and the data controller. Article 7 puts this as "processing [which] is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract".

3.1.3 Legal obligation

Even when the data subject has not given their consent or signed a contract, there are valid legal grounds for data processing. The most straight-forward of these legal grounds is a legal obligation, when (as Article 7 states) "processing is necessary for compliance with a legal obligation to which the controller is subject".

3.1.4 Vital interests

A fourth legal ground for data processing applies when "processing is necessary in order to protect the vital interests of the data subject". These *vital interests* are not further specified by Article 7, but the Article 29 Working Party comes to the rescue. Vital interests are defined by the Article 29 Working Party in Working Document 01/2012 as a case of emergency: "the processing must relate to essential individual interests of the data subject or of another person and it must in the medical context be necessary for a life-saving treatment in a situation where the data subject is not able to express his intentions." In other words: vital interests relate to cases where the processing of data is necessary

for the well-being of the data subject, another person or the public in general, in a situation where the data subject is not able to express their intentions. An obvious example would be a situation where a doctor needs the bloodtype of the patient on their surgery table, while the patient is unconscious and unable to share their data.

3.1.5 Public interest

Public interest can be another legal ground for data processing. Article 7 of the DPD defines this as when "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed". Public interests nclude public health, social protection and/or economic well-being.

3.1.6 Legitimate interests

The last possible legal ground is based on legitimate interests of the data controller, when "processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1)". Since Article 7 does not further elaborate or define *legitimate interests*, we once again turn to the Article 29 Working Party. Legitimate interests are mentioned in Opinion 08/2012, where they are defined as "a legal ground for processing personal data, if, and to the extent in which, certain conditions have been fulfilled, requiring a balancing test to be executed, in the light of the circumstances of each case." Note the last sentence of 7(f): processing of personal data is allowed for the purposes of legimitate interests, except where such interests are overridden by fundamental rights. It refers to Article 1 (1) of the DPD, which states that "Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data." The Article 29 Working Party further elaborates on legitimate interests in Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC" and states that "the legitimate interests of the controller (or third parties) must be balanced against the interests or fundamental rights and freedoms of the data subject. The outcome of the balancing test largely determines whether Article 7(f) may be relied upon as a legal ground for processing." Briefly summarized, they propose such a balancing test by defining the legitimate interests of both the data controller and the data subject. According to the Article 29 Working Party the legitimate interest of the data controller must be:

- in accordance with applicable EU and national law,
- "sufficiently clearly articulated to allow the balancing test to be carried out against the interests and fundamental rights of the data subject" and
- represent a real (as opposed to speculative) and present interest.

On the other side of the scale lie the interests and the fundamental rights of the data subject. In their aforementioned opinion, the Article 29 Working Party even omits the adjective "legitimate" when talking about the interests or rights

of the data subject, intentionally implying a much wider range of interests than in the case of the data controller. In other words: whereas the data controller must always specify the purpose of processing, the data subject is not required to specify why they would not wish to have their data processed. This relates to the right to object to profiling ("The data subject should have the right to object to the processing, on grounds relating to their particular situation and free of charge") as stated in Recital 38 of the pGDPR. The balance can then be provisionally made up by weighing the legitimate interests of the data controller against the impact on the data subject. The Working Party adds the need for additional safeguards applied by the data controller: "[t]he more significant the impact on the data subject, the more attention should be given to relevant safeguards."

One relevant example of such balance was made in the Google v Spain case[7]. A Spanish man whose home had been auctioned years ago complained that, although his financial problems had all been resolved by now, a reference to this still appeared in a newspaper archive and in Google's search results. The court then ruled that "even initially lawful processing of accurate data may, in the course of time, become incompatible with the directive where those data are no longer necessary in the light of the purposes for which they were collected or processed. That is so in particular where they appear to be inadequate, irrelevant or no longer relevant, or excessive in relation to those purposes and in the light of the time that has elapsed" (paragraph 93 of the ruling). This is where the court made a very relevant decision regarding balancing legitimate interests of the data controller against the fundamental rights of the data subject: paragraph 97 of the ruling states that these fundamental rights of the data subject "override, as a rule, not only the economic interest of the operator of the search engine but also the interest of the general public in finding that information upon a search relating to the data subjects name". Note how this leads to a possible problem with PDMs: a Personal Data Management tool with a business case mainly focusing on making a profit would automatically make them a data controller with an economic interest, implying they are not allowed to control (process, but also store) personal data that are irrelevant, inadequate et cetera.

3.2 Purpose limitation

Article 6(1)(b) of the DPD states that Member States, after they have guaranteed the personal data are processed fairly and lawfully (6(1)(a)), shall provide that personal data must be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes." As with consent, this requires more specific definitions of terms like *specified*, *explicit* and *legimitate*. The Article 29 Working Party discusses these in Opinion 03/2013 on purpose limitation:

specified implies that the data controller must "carefully consider what purpose or purposes the personal data will be used for, and must not collect

[7]Google v Spain - `http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf` and `http://curia.europa.eu/juris/document/document.jsf;jsessionid=9ea7d2dc30d59eba99abf29b4349b30a8b9912962cdf.e34KaxiLc3qMb40Rch0SaxuOchj0?text=&docid=152065&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=153961`

personal data which are not necessary, adequate or relevant for the purpose or purposes which are intended to be served." This also means that personal data can not be used for incompatitble purposes than initially collected for.

explicit means that the purposes of data collection must be "clearly revealed, explained or expressed in some intelligible form."

legitimate has a broader meaning than just the legal grounds for data processing as listed in Article 7. Besides meeting at least one of those criteria, legitimate also means "that the purposes must be in accordance with all provisions of applicable data protection law, as well as other applicable laws such as employment law, contract law, consumer protection law, and so on." Considering data processing, legitimate also often relates to Article 8 of the European Convention of Human Rights. So, legimitate here translates to "in accordance with the law" in general.

Furthermore, Article 6(1)(c) to (e) should also be satisfied. This implies that personal data must be:

- adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;
- accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified;
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed.

There are some exceptions. Article 6(1)(b) also states that "further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards." Those *appropriate safeguards* seem pretty vague and loosely defined. According to the Article 29 Working Party "it is up to each Member State to specify what safeguards may be considered as appropriate." Since those *historical*, *statistical* or *scientific purposes* also seem loosely interpretable, the Working Party suggests "different kinds of safeguards, including technical and organisational measures for functional separation, such as full or partial anonymisation, pseudonymisation, aggregation of data, and privacy-enhancing technologies" (from Opinion 03/2013 on purpose limitation). If personal data are further processed for different compatible purposes than originally collected for, the requirements described above should also hold for those new purposes, or the data must be either recollected, anonymised or deleted. The latter is best known as the *right to be forgotten*, made famous by the Court of Justice of the European Union when a Spanish citizen requested Google to delete his no longer adequate or relevant data from their search results. The court ruled that the right to be forgotten applies "where the information is **inaccurate, inadequate, irrelevant or excessive** for the purposes of the data processing"[8], but at the

[8]Factsheet on the "Right to be Forgotten" Ruling - `http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf`

same time "explicitly clarified that the right to be forgotten is not absolute but will always need to be balanced against other fundamental rights, such as the freedom of expression and of the media". These rulings relate to Article 12 and Article 14 of the DPD, which state that Member States shall grant the data subject the right to

- obtain from the controller "as appropriate the rectification, erasure or blocking of data the processing of which does not comply with the provisions of this Directive, in particular because of the incomplete or inaccurate nature of the data" (Article 12(b)), and
- "at least in the cases referred to in Article 7 (e) and (f), to object at any time on compelling legitimate grounds relating to his particular situation to the processing of data relating to him, save where otherwise provided by national legislation. Where there is a justified objection, the processing instigated by the controller may no longer involve those data." (Article 14(a)).

This implies that the applicability of the right to be forgotten has to be determined per single case, taking into consideration the type and sensitivity of the information and the public interest of the accessibility to that information.

3.3 Decision making based on automated processing

Article 15.1 of the DPD states that the data subject always has the right to object to a decision "which produces legal effects concerning him [the data subject] or significantly affects him and which is based solely on automated processing of data". Furthermore, those decisions based on automated processing are only allowed in certain scenarios, which are defined in Article 15.2:

- The decision was made according to a contract (15.2a), or
- the decision was authorized by a law (15.2b).

Under the current version of the pGDPR, the consent of the data subject is a third condition that allows for profiling.

In all cases, the Article insists that suitable measures are taken to safeguard the data subject's legitimate interests, but does not further elaborate on those "suitable measures" other than "arrangements allowing him to put his point of view" (15.2a). Article 12 gives some explanation, stating that the data subject always has the right to obtain the following information from the data controller:

- confirmation as to whether or not data relating to them are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed,
- communication to them in an intelligible form of the data and of any available information as to their source,
- knowledge of the logic involved in any case of automatic processing of data. This logic should also be presented to the data subject in an intelligible and understandable form, and is especially interesting when it comes to observed and inferred data.

Furthermore, the Article 29 Working Party issued a Recommendation (Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware) on the protection of individuals with regard to the processing of personal data, expressing their concern "about all kinds of processing operations which are presently being performed by software and hardware on the Internet without the knowledge of the person concerned and hence are "invisible" to him/her." This Recommendation focuses on internet hard- and software (mainly browsers) that collect and process user data, and thus is not applicable to automated decisions causing legal effects as described earlier. It does say something about suitable measures: apart from the points already mentioned in the DPD, the Article 29 Working Party also suggests that data controllers should "give the capacity to the data user to easily access any data collected about him/her at any later stage."

3.4 Pseudonymous Data & Profiling

In the previous sections I have briefly shown the laws en requirements regarding the processing of personal data. The principles of data protection as described in the DPD and the pGDPR do not apply to anonymous data, meaning data that are not related to any *identifiable natural person*. A subset of personal data that still has similarities with anonymous data is *pseudonymous data*. Article 4(3)(b) of the pGPDR defines pseudonymisation as:

> "'pseudonymisation' means the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person"

In their Advice on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation, the Article 29 Working Party defines *profiling* as:

> "any form of automated processing of personal data, intended to analyse or predict the personality or certain personal aspects relating to a natural person, in particular the analysis and prediction of the persons health, economic situation, performance at work, personal preferences or interests, reliability or behaviour, location or movements."

Unlike automated processing of personal data, profiling based solely on the processing of pseudonymous data is not presumed to be significantly affecting the interests, rights or freedoms of the data subject[9]. Yet, and this is particularly interesting for some Personal Data Management systems I will discuss later, as soon as this profiling leads to the possible identification of a data subject, or the data controller is able to attribute (some of) the pseudonymous data to a specific data subject, the data are no longer pseudonymous. To underline the trickiness of pseudonymity once more, recent study by [de Montjoye et al., 2013]

[9]Recital 58(a) of the pGDPR

on inferred mobile data showed that "four spatio-temporal points are enough to uniquely identify 95% of the individuals". Earlier, in 2009, [Ohm, 2010] already wrote that pseudonymous data can often easily be deanonymized partially or entirely, and that individuals can be reidentified.

3.5 Summary

The following checklist summarizes the legal issues discussed above, and can be used when testing Personal Data Management systems for their compatibility with those legal requirements:

1. Has the data subject given their consent according to Articles 7(a), 8(2)(a) and 2(h), or
2. is processing necessary because of a contract (7(b)), or
3. is processing necessary because of a legal obligation (7(c)), or
4. is processing necessary because of protecting vital interests of the data subject (7(d)), or
5. is processing necessary because of public interests (7(e)), or
6. is processing necessary for the legitimate interests of the data controller or any third party to whom data are disclosed and are those legitimate interests not overridden by the fundamental rights and freedoms of the data subject (7(f))?

If the answer to (at least) one of these questions is "yes", the following questions to be checked are:

1. Are the data collected for specified, explicit and legitimate purposes (6(1)(b)), and
2. are the data either processed further in a way compatible with these purposes or processed further for historical, statistical or scientific purposes (6(1)(b)), and
3. are the data adequate, relevant, not excessive in relation to purposes, accurate, up to date and kept in a form permitting identification no longer than necessary (6(1)(c-e)), and
4. has the data controller provided the data subject with its identity, processing purposes and any further necessary information (10/11)?

If the answer to *all* these questions is "yes", personal data are processed lawfully according to Articles 6, 7, 10 and 11 of the DPD. The last checklist is related to automated processing (see section 3.3):

1. Was the decision made according to a contract (15(2)(a)), authorized by a law (15(2)(b)) or based on consent of the data subject (15(2)(c)), and
2. have suitable measures been taken to safeguard the data subject's legitimate interests (15(2)), and

3. does the subject have knowledge of the logic involved in the processing of their data (12(c)), and

4. was this knowledge presented to the data subject in an intelligible and understandable form (12(c))?

If the answer to *all* these questions is "yes", the automated processing of data was lawful according to Article 12 and 15 of the DPD.

4 Personal Data Management systems

4.1 Mydex

One example of a PDM is the Mydex project by [Heath et al., 2013]. They introduce a Personal Data Store, which they define as "a service for individuals that helps them collect, store, manage, use and share their own personal data for their own purposes"[10]. The main idea behind this PDS is to give the individual full control over their personal data. The entire Mydex environment runs in beta at the moment of writing, but it already allows developers and individuals to test and try the system.

4.1.1 Business case

Mydex has been set up as a so-called *Community Interest Company* or simply *CIC*. This means for one thing that Mydex in fact is a commercial organization which can make profits, but always has to serve a higher social benefit. More specifically this means that Mydex has to (re)invest 65% of their profits in a social cause, in this case to help individuals realise the value of their personal data. The remaining 35% can then be distributed over their shareholders. Mydex makes money by charging organizations a fee for using their data sharing features. A second requirement of CICs is to create an asset lock, or "a legal promise stating that the companys assets will only be used for its social objectives"[11]. For Mydex this implies they forever have to stick to their chosen social purpose, and that Mydex as a company can never be sold to another organization with different purposes for their assets.

4.1.2 Actors

4.1.2.1 Individuals Within Mydex, the **data subject** is called the *individual*. The main reason for shifting away from the more common division into data subject, data controller and data processor is that with the introduction of the personal data store, the individual is no longer only the data subject. Having complete control of their own personal data, individuals play the role of data controller and subject simultaneously. They can give or revoke consent to sharing their data with other entities at any time. Furthermore, all data and communication channels are encrypted with their own private key, implying that none of the organisations, applications or the Mydex system has means of accessing the data either. The user can choose their own location to store their Personal Data Store in, both locally or in another data centre. The Mydex architecture supports the user in moving their PDS to another data service, or storing a copy elsewhere.

4.1.2.2 Organisations Before using Mydex, all connecting organisations and applications have to identify themselves to the framework. All data sharing within Mydex is subject to a data sharing agreement between the individual and the organisation, allowing both the data subject and Mydex themselves to

[10] Mydex FAQ - `https://mydex.org/faqs/`
[11] Setting up a social enterprise - `https://www.gov.uk/set-up-a-social-enterprise`

immediately take action in the event of any breach of contract. All connecting organisations are **data controllers** in legal terms.

4.1.2.3 Mydex Mydex itself is the facility which enables (or disables) data sharing and storage. Organisations have to sign a contract with Mydex as well as pay a small fee for accessing their services. The latter is only interesting for commercial purposes, but the former allows Mydex to take action whenever a connecting organisation breaks a contract, since this would automatically be a breach of their terms of connection. Apart from that, Mydex is only facilitating the service and claims it cannot access, view or edit any of the stored data. This should be enforced technically by using private keys only known to the individual. As with every PDM, Mydex is also a **data controller**: see also section 2.1.1.

4.1.3 Framework

In this paragraph I will take a closer look at the framework of Mydex. I will do so by using the approach as suggested by [Bus and Nguyen, 2013] and divide the Mydex framework into infrastructure, data management and user interaction as introduced in chapter 2.1.

4.1.3.1 Infrastructure The following information is all derived from [Heath et al., 2013], the Mydex developers environment[12], the Mydex FAQ[13] and a (possibly self-written) interview[14]. The Mydex environment is currently only cloud based, making the entire infrastructure highly similar to a social platform, with the future possibility for individuals to store their entire datasets locally. At the time of writing it remains unclear how Mydex will make this possible.

Encryption is said to be used on the data, although technical details (for example encryption schemes) are not provided. Although security by design is one of the legal requirements, I will assume that - within the scope of this thesis - the encryption of the data works fine and is indeed secure. I would however strongly suggest Mydex to use open source software in order to gain trust from their (future) users. The same goes for authentication, where Mydex intends to extend the security through biometrical authentication techniques in the future.

More related to purpose limitation, Mydex uses metadata - on top of the personal data - describing the content and the related permissions of the individual. Individuals can edit or withdraw these permissions at any given time within their personal data store. Such an additional metadata layer however seems not technically capable of destroying the data underneath in the event of changing permissions.

4.1.3.2 Data Management One of the key aspects of Mydex is the handling of data. All data stores are both encrypted and distributed over several servers, ensuring there is no single central (hence vulnerable) database with all the data. At the time of writing all data are stored in the cloud, with - as mentioned before - the future possibility for indiviuals to store their entire datasets locally.

[12]Mydex Developer Community - `https://dev.mydex.org`
[13]Mydex FAQs - `https://mydex.org/faqs/`
[14]Mydex Tech Documentation - `http://hub.personaldataecosystem.org/wagn/Mydex`

The Mydex Tech Documentation[15] mentions the feature to export all data into a single "file format of which includes the meta data describing the content so it is machine readable", although more specific details about this file format or (meta)data structure is lacking. Mydex itself can only see these metadata, implying it does not have reading access to the personal data of the individuals.

As mentioned earlier, all data controllers have to sign a contract with Mydex. These data controllers, referred to as both *organisations* and *third parties*, get verified and certified by Mydex to use the environment after paying a small fee. This ensures that in the case of breach of contract, hence unlawfully processing user data, these data controllers are responsible and liable. Unfortunately, I was not able and/or allowed to view these contracts myself.

Users and organisations sign a Data Sharing Agreement between them, providing specific information about what data are to be shared in either direction, and what specific rights the organisation are being granted. These specifications include both the purposes and limitations for using the data, like use cases the organisations can use the data for and/or the number of times the data can be accessed.

4.1.3.3 User Interaction As summarized in section 2.1, [Bus and Nguyen, 2013] suggest that each personal data management system should offer simple and intuitive tools for the user "to have meaningful interaction with service providers regarding the permissions and policies associated with the use of their personal data." The social platform of Mydex offers the individual two possible ways for entering personal data into the system, roughly corresponding with the distinction between volunteered and observed data (see section 1.1). The first is the most intuitive: after obtaining the necessary credentials for using the platform, the individual either uploads or manually enters their data into forms, divided into several (user managable) tabs. See figure 1 for an example. The individual can then select which organizations they want to share some of their data with, and further specify which data are shared with whom in a similar tabs-and-forms-environment (see figure 2). The individual also has the choice to make their profile either publicly visible or not, the first allowing third parties to request data connections with the individual under their own specified conditions. Note that all of this relates to volunteered data only: third parties will still be interested in browsing behaviors, phone call histories and other observed and inferred data outside the Mydex environment.

Mydex also offers a second, more automatic feature of acquiring data. The individual can choose to synchronise certain information with their personal data store automatically using "a range of import, transform and load utilities as well as establishing, one time, one way connections right [through to] persistent bidirectional connections for enduring data sharing and updates e.g. credit card history, mobile phone call history, digital receipts, bank transactions, utility energy consumption data, loyalty card transactions" (source: Mydex Tech Documentation[16]). By constantly sending out these updates on phone call history, bank transactions and such, this second approach can be seen as a combination of both volunteered (because the individual has to enable it) and observed data.

[15] Mydex Tech Documentation - `http://hub.personaldataecosystem.org/wagn/Mydex`
[16] Mydex Tech Documentation - `http://hub.personaldataecosystem.org/wagn/Mydex`

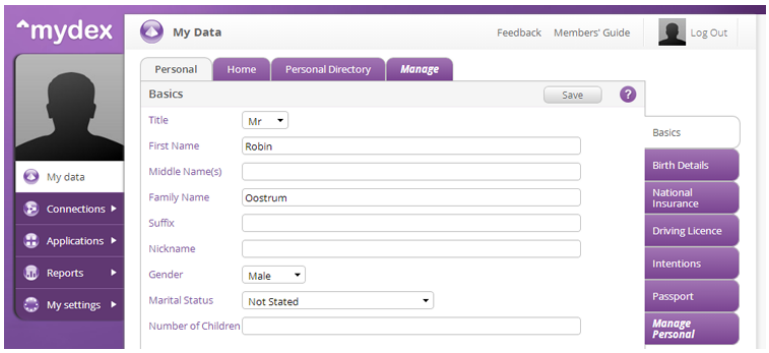

Figure 1: Example data fields in Mydex

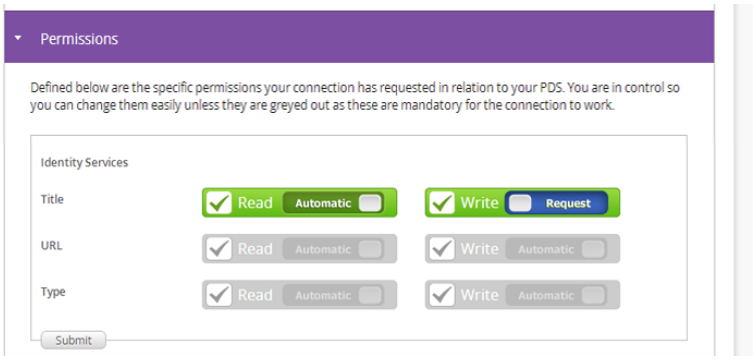

Figure 2: Permission handling in Mydex

4.1.3.4 Data types When we look at the different types of data that are possibly involved in data processing, Mydex mainly focuses on volunteered data like age, name, address, credit card details and declared energy use. Observed data like realtime energy use and telephone data can be uploaded automatically to the data store, but - at the moment - inferred data (web analytics for example) are out of the scope of the Mydex environment. Data are not being anonymised or pseudonymised (see figure 1 for example), neither does Mydex significantly focus on sensitive (health) data.

4.1.4 Compatibility with Legal Requirements

4.1.4.1 Consent As we have seen, Mydex allows the user to specify which data are shared with whom. I will now take a more detailed look at the legal definition of consent (see section 3.1.1) to see whether Mydex really adds something on consent or not.

Freely Given According to the definition of freely given consent as stated by the Article 29 Working Party, "there must be no risk of deception, intimidation or significant negative consequences for the data subject if he/she does not consent" (from Opinion 15/2011 Consent). Within Mydex, consent can be given per attribute: the individual decides which organisations are allowed access - and

which are not (see also the paragraph on user interaction (section 4.1.3.3) and figure 2). Regarding the addition from Article 7.4 of the pGPDR (on performing contracts and offering services even when the data subject refuses consent for the collection of irrelevant and unnecessary data), Mydex allows individuals to allow or deny permissions one by one. In accordance with this addition, the Mydex connection between an organization and the individual persists even if the latter should decide to deny all possible data sharing permissions. However, Mydex cannot guarantee that an organization continues offering their service after an individual withdraws or denies their consent. Furthermore, organizations can still force individuals to share their personal data (always consider the gun-against-head-scenario): Mydex cannot verify whether given consent is completely freely given or not.

Specific "Blanket consent without determination of the exact purposes" (Article 7 of the DPD) is not valid. Applying this to Mydex, the specific purposes of the revelation of personal data must be known to the individual when giving their consent to a specific organisation. Taking a closer look at the *specific* requirement, the Working Party adds that the given consent "cannot apply to an open-ended set of processing activities. This means in other words that the context in which consent applies is limited." Within Mydex, as we have seen in section 4.1.3.2, users sign Data Sharing Agreements with organisations to specify these purposes (see also the 'Contract' section in 4.1.4.2). However, and this will come back several times in the following paragraphs: once the data are *out* of the Mydex enviroment (for example by screen capturing or copy-pasting decrypted data), the individual loses control. Basically this means, regarding the *specific* requirement, that organisations can do whatever they want with the data once they obtained access.

Informed What we have seen above is also in accordance with the informed requirement stating that the organisations must inform the individual with the identity of the controller and the purposes of the processing, as well as any possible further information such as the recipients of the data (see section 3.1.1). Other information such as the identity of the concerning organisation and the recipients of the data are always known to the individual since they will always have given consent themselves, and organisations are both verified and bound to a legal contract with Mydex. However, it remains out of the scope of Mydex what happens when the personal data are out of the Mydex environment: for all we know the concerning organisation keeps an offline copy of the data, sharing it with third and fourth parties. While this is obviously illegal, it seems hard (if not impossible) to detect such fraud. Should fraud be detected however, the third parties responsible will be held liable for breaking the contract they signed with Mydex.

Unambiguous As seen in section 3.1, unambiguous consent is defined by the Article 29 Working Party as consent "that leaves no doubt as to the individual's intention to provide consent." Mydex enforces unambigious consent by requiring the individual to literally tick a box in order to give consent.

Explicit Mydex also ensures that given consent is always explicit. By ticking the box for giving consent, the individual explicitly expresses "his/her wish to have his/her data processed for certain purposes" (Article 8.2 of the DPD).

4.1.4.2 Contract One other possible ground for data processing is the performance of a contract to which the data subject is bound. This means that if there is a contract between an individual and an organization, which allows the latter to process personal data for the fulfillment of that contract, Mydex should facilitate this. Within Mydex the data subject is bound to a contract with Mydex itself, a Data Sharing Agreement, obliging the individual by their terms and conditions to release their relevant data if a contract between individual and organisation is presented. This also means that a contract between an individual and an organisation basically obliges the individual to grant permission for sharing the relevant data.

A different solution for Mydex would be to implement a way to share personal data *without* asking the individual for their permission, in cases where other legal grounds apply: Mydex would then still function as a Personal Data Management system by showing the user which attributes are shared with whom, based on what legal requirement. This would not only automize the *contract* requirement, but also solve issues with other legal grounds like legal obligation and vital interests, as can be seen below. Note that this would imply that Mydex is able to decide whether a data sharing request is valid for one of these grounds. At the moment it is not possible to collect data from a members personal data store without user action.

4.1.4.3 Legal obligation As with other (both on- and offline) dataholders, a legislative body or government can oblige the individual to release their data if there is a statutory obligation. In Mydex, this means the individual is obliged to enter their credentials and agree to share their data with the data controller. Note that releasing data in this context is significantly different than giving consent.

4.1.4.4 Vital interests At first glance it seems impossible for the doctor to get the necessary data: the patient has to enter their credentials and agree to share relevant data with the doctor. Mydex could solve this by adding a kind of "emergency permission" to the metadata of relevant health data, allowing doctors and others access in the case of such an emergency.

4.1.4.5 Public interest As with vital interests, it seems hard for Mydex to guarantee that public interest can be a ground for data processing within their environment. Ideally, like suggested above in paragraph 4.1.4.4, permissions will be added in the form of metadata that allow specific data controllers to access the data if the public interest requires so. The question then remains who determines when a request to process personal data is necessary for the public interest, and how that data may then be processed. Mydex currently does not implement this.

4.1.4.6 Legitimate interests In the case of processing personal data for the purposes of *legimitate interests* of an organization or third party, after a thorough balancing test has shown that indeed those legitimate interests are not overridden by the interests and fundamental rights of the individual, the latter should always have the right to object against this decision. Mydex at least guarantees that:

- only the concerning organization can see the data (as discussed before, Mydex is unable to view actual contents of the personal data stores, and data are only shared with the concerning organization), and
- the user gets notified that an organization processed their personal data (since the user has to enter their credentials and share the data themselves).

The question is who performs the balancing test and determines what data are relevant for sharing for the legitimate interests of the data controller. In a future situation, an organization enters their interests and demands into Mydex, which then automatically weighs these against the interests and fundamental rights of the individual, and (only) shares the relevant personal data with the organization involved. At the time of writing, there are no automated tools available yet for making such a judgment.

4.1.4.7 Purpose Limitation We have seen how Mydex deals with consent, contracts and other possible legal grounds for data processing, but how does it deal with purpose limitation? In this paragraph I will take a closer look at the requirement (as stated in 3.2) that personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.

As described in section 4.1.3.3, the individual selects which attributes they share. Mydex requires organisations to specify the goals and purposes of this data sharing (in the Data Sharing Agreement, see section 4.1.3.2). Looking at the DPD requirements for purpose limitation (section 3.2), this only partially meets the requirements: although organisations would break the contract with the PDM and can be held accountable, Mydex cannot technically guarantee that personal data are not used for other purposes than specified by the organization.

Even when data are in fact being processed for their specified goals, purpose limitation is not guaranteed: if personal data are further processed for incompatible purposes than originally collected for, the requirements described above should also hold for those new purposes, or the data must be either recollected, anonymised or deleted. As we have seen in the paragraphs above, Mydex cannot guarantee that third parties do not store acquired data outside the secure Mydex environment and thus use personal data for purposes other than initially collected for. As mentioned before, this is the biggest issue with personal data management systems using a secure environment, like Mydex.

4.1.5 Conclusion

A summary of the security and legal requirements of Mydex can be found below in table 1. In the sections above, I have shown that Mydex can offer some improvements regarding data processing. For instance, the data subject is able

Feature	Supported by Mydex?
Data minimization	✓
End-to-end encryption	✓
Meaningful consent	✓
- Explicit consent	✓
- Specific consent	✓
Other grounds	
- Contract	✓
- Legal obligation	✓
- Vital interests	
- Public interests	
- Legitimate interests	
Purpose specification	✓
Purpose limitation	
Anonymization	
Pseudonymisation	
Legitimate profiling	
User can terminate data processing	✓

Table 1: Mydex round-up

to select the attributes they share and with whom, meaningful consent is ensured by requiring user action and the Data Sharing Agreement, and all personal data are encrypted end-to-end.

But we have also seen (for example in section 4.1.4.1) how Mydex cannot guarantee that organizations comply with some legal requirements from the DPD: Mydex cannot guarantee that an organization continues offering their service after an individual withdraws or denies their consent (Article 7.4 of the pGPDR) and Mydex cannot guarantee that third parties do not store acquired data outside the secure Mydex environment and thus use those for unlawful or unspecified purposes. Also, the fact that Mydex always requires user action before data are being shared makes it hard to process data on other grounds than consent or contract.

4.2 IRMA

A second example of a Personal Data Management system is *I Reveal My Attributes*[17] (from now on referred to as *IRMA*), developed by the digital security department of the Radboud University. The main idea behind IRMA is to not reveal more personal data than necessary. When ordering a beer in a pub, it should be sufficient to prove you are at least 18 years (or 16 or 21) old, without revealing personal attributes like your name, nationality or even your exact date of birth. Only showing relevant (non-identifying) attributes prevents the possible linking of multiple actions. This in contrast to a telephone smart card for example, which has a unique identifier and allows the phone provider to link multiple actions (phone calls, text messages et cetera) and trace them back to the user. Unlinkability and untraceability make it impossible to profile users of IRMA, while using non-identifying attributes like age or 'senior citizen' does not even reveal one's identity.

4.2.1 Business case

IRMA is a cooperation between the following four organisations:

- Radboud University Nijmegen
- SURFnet
- TNO
- SIDN

As opposed to Mydex, IRMA is primarily a research project. In other words: making a profit is not the main goal of IRMA. The parties mentioned above all invest in the project by funding technology like smart cards, readers and such. If the test phase is successful, IRMA will be looking for more involved parties in the future. The ultimate goal is to transfer the entire technology and infrastructure to an independent foundation.

4.2.2 Actors

4.2.2.1 Individuals Each individual user has their own personal *IRMA card*: a smart card with their picture on the outside, protected by a personally chosen PIN only the user should know. The user then can obtain *attributes* from an *Attribute Issuer* by downloading them to the card, after the user authenticated themselves to the Attribute Issuer. Examples of such attributes are[18]:

- I am a student
- My nationality is...
- My address is...
- I am an inhabitant of the city of...

[17] I Reveal My Attributes - `https://www.irmacard.org/irma`
[18] source: `https://www.irmacard.org/irma`

- I have a valid public transport ticket

Those attributes can then be used in various transactions with so-called *Relying Parties*. The Relying Party shows the cardholder which attributes it will check, who confirms this by entering their PIN to prove - along with the picture on the outside of the card - ownership of the IRMA card.

The invidual is responsible for keeping their attributes up-to-date: each attribute contains a validity date after which it expires and cannot longer be used. By returning to the original issuer, the individual can obtain a new attribute or extend the validity of the one issued before. They have full control over their personal data by choosing which attributes get shared with every Relying Party, provided that the Relying Party can only see the attributes it claims to check. The latter has to be technically enforced.

4.2.2.2 Attribute Issuer The Attribute Issuer allows IRMA card users to download attributes onto their personal cards. Therefore, Attribute Issuers need to be trusted authorities in the area of the provided attributes. Example Attribute Issuers include:

- Universities
- National and local authorities
- Banks and insurance companies
- Internet service providers

An Attribute Issuer lets the user authenticate themselves before showing them all the available attributes. Hence, the main responsibility of the Attribute Issuer is to ensure the attributes it issues are valid. This raises the question how IRMA deals with attributes that are no longer valid: issued attributes should be validated by Attribute Issuers every now and then, and temporary attributes (like age, license suspensions etc.) should automatically become invalid using time stamps or such. I will further discuss this in the technical section of this capter (see: paragraph 4.2.3).

4.2.2.3 Relying Party Relying Parties are the parties offering any kind of (online or offline) service or product and rely on the claims made by the Attribute Issuers and which are shown by the individual users. The Relying Party shows the individual which attributes it wants to check: the cardholder confirms this by authenticating themselves both by PIN and photo (only in offline authentication) on the IRMA card, allowing the Relying Party to cryptographically (see: paragraph 4.2.3) check those attributes. Obviously it should be technically infeasible for the Relying Party to see other attributes than allowed.

4.2.2.4 Scheme Manager The Scheme Manager is an independent party overlooking the network of relying parties, issuers and individuals. It sets the rules for each party and controls the software and architecture of the IRMA Card. Before using IRMA, Relying Parties sign a contract with the Scheme Manager to commit themselves to behave as promised. They then receive a revokable certificate that gets checked by the card during the transaction, which

contains the attributes the specific Relying Party is allowed to see. At the time of writing, IRMA is still running in a test phase without a Scheme Manager as there are only a couple of Relying Parties involved.

4.2.3 Framework

In this paragraph I will take a closer look at the framework of IRMA. I will do so by using the approach as suggested by [Bus and Nguyen, 2013] and divide the IRMA framework parts into infrastructure, data management and user interaction as introduced in chapter 2.1.

4.2.3.1 Infrastructure IRMA uses a mixture of IBM's Idemix ([Camenisch and Van Herreweghen, 2002]) and the Microsoft U-Prove ([Brands and Paquin, 2010]) frameworks, which are both open technologies that combine several attributes into one credential and allow *selective disclosure* of attributes: this allows the selection of only a subset of the attributes within such a credential. Each credential is bound to the card using the private key of the card, ensuring a credential cannot be transferred from one IRMA card to another. This private key works with a PIN chosen by the individual owner of the IRMA card, and ensures ownership in online transactions. (For offline transactions, the photo on the outside of the card would be sufficient. The PIN could be seen as an extra safety measure.) The set-up for obtaining such credentials from an Attribute Issuer is described by [Vullers and Alpár, 2013]: "First the user authenticates to the issuer in some reliable but unspecified manner (which may be face-to-face). Once the authentication succeeds, the issuer collects attributes for this user from trusted databases. The user and issuer then carry out a cryptographic protocol in which the attributes are combined into a credential signed by the issuer. The resulting credential contains attributes concerning the user and also their secret key." Note that the credential is signed by the Attribute Issuer: this ensures both the *authenticity* and the *integrity* of the credential, meaning the Attribute Issuer claims the attributes in the credential hold for the individual at that moment and meaning that these attributes have not been altered since they were issued. To ensure the validity of a credential at a specific moment in time, each credential also features its own expiry date.

Between the individual and the Relying Party a secure communication channel is always required, since the disclosure of attributes to any third party is undesirable. The technique behind this secure channel is described by [Alpár and Hoepman, 2013], who propose a secure Attribute-Based Credential session protocol called *Implicit Card Authentication* or *ICA*. Setting up this secure channel, the Relying Party authenticates itself to the IRMA Card (and its owner). The protocol is summarized in figure 4 (from: [Alpár and Hoepman, 2013]):

Where C denotes the individual's IRMA Card, V the Verifier or Relying Party, n a nonce, k the session key and pkv and skv respectively the public and secret key of the Relying Party. Note that the Relying Party (or *Verifier* in the protocol) authenticates itself to the card as it requires its private key skv to decrypt nonce n_c after step 2. The secure channel subsequently guarantees the confidentiality and integrity of the attributes shown in the transaction.

The IRMA architecture explicitly chooses the use of smart cards over mobile devices. Modern smart cards provide secure use of private keys and attributes, whereas tablets, mobile phones and such are susceptible to malware. Smart

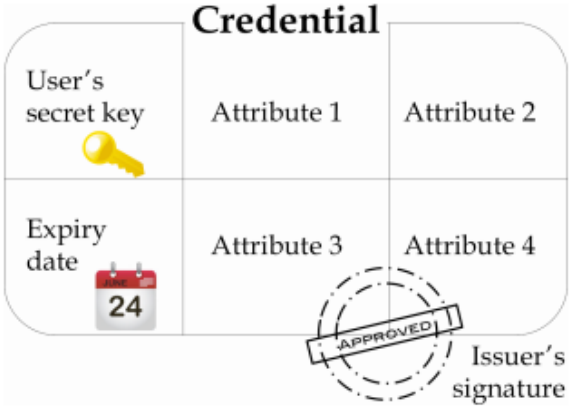

Figure 3: The structure of a credential ([Alpár and Jacobs, 2013])

cards are designed (through both soft- and hardware) to guarantee that those private keys are never sent over untrusted channels. The use of mobile phones - by implementing IRMA into SIM cards or using alternative phone compatible smart cards - is rejected by IRMA developers since mobile phones get lost/stolen, people often change their phone and many people use a phone that is provided by their employer[19].

4.2.3.2 Data Management Data minimisation is a key aspect of the IRMA card architecture. In fact, there is no third party storing any data about individuals: all the attributes are on the smart card of the user themselves. As the IRMA website[20] puts it: "When you wish to prove to a shop keeper that you are over 18, you can do so directly, without going through some central infrastructure that can keep track of what your attributes are [or] who you are showing them to."

4.2.3.3 User Interaction As summarized in section 2.1, [Bus and Nguyen, 2013] suggest that each personal data management system should offer simple and intuitive tools for the user "to have meaningful interaction with service providers regarding the permissions and policies associated with the use of their personal data." Applying this to IRMA, we have already seen that the interaction between user and Relying Party is pretty straight-forward and intuitive: the Relying Party shows the user which attributes it wants to check, whereupon the user enters their PIN to confirm the transaction of personal data.

The biggest responsibility for the user is to keep these attributes up-to-date by interacting with the Attribute Issuer. This interaction, although less related to permissions and policies, is also in accordance with the proposed "meaningful interaction" by [Bus and Nguyen, 2013]. After authenticating to the Attribute Issuer, the user picks their desired attributes from the database (of the Attribute Issuer) and downloads them to their card by simply ticking

[19]Why are smart cards used for IRMA? - `https://www.irmacard.org/irma/#05`
[20]I Reveal My Attributes - `https://www.irmacard.org/irma`

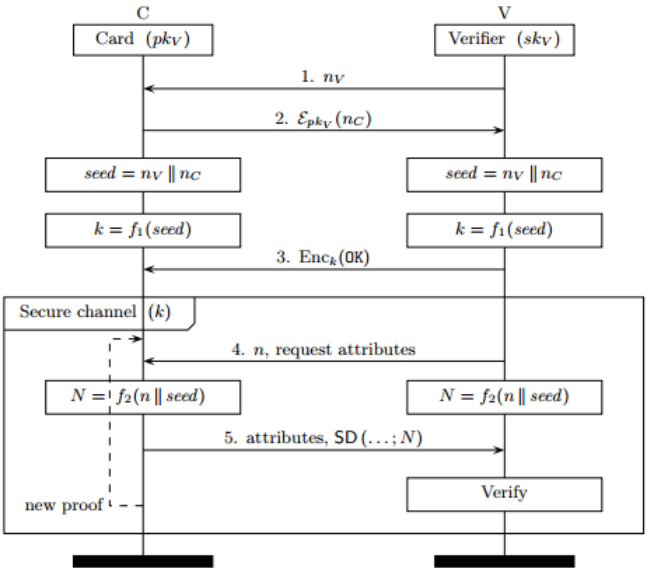

Figure 4: ICA Protocol

boxes in a user-friendly web interface.

4.2.3.4 Data types When we look at the different types of data that are possibly involved in data processing, IRMA mainly focuses on volunteered data like age, name, address, credit card details and declared energy use. The architecture of IRMA makes it harder to share observed data like web surfing behaviour or realtime energy use, since attributes always have to be downloaded to the smartcard first. This also implies that IRMA does not deal with inferred data.

Within the IRMA environment, anonymous attributes (like "allowed to drink" or "licensed to drive a car") remain anonymous to the Relying Party, unless an additional attribute is presented with which the data subject can be identified. As we have seen in section 3.4, this type of data is defined as *pseudonymous* data.

4.2.4 Compatibility with Legal Requirements

4.2.4.1 Legal Grounds

4.2.4.1.1 Consent As we have seen, attributes may only be processed after the Relying Party shows which attributes it wants to check and the user has subsequently given their consent by entering a PIN. I will now take a more detailed look at the legal definition of consent (see section 3.1.1) to see whether IRMA really adds something on consent or not.

Freely Given Quoting the Article 29 Working Party once more, "there must be no risk of deception, intimidation or significant negative consequences

for the data subject if he/she does not consent" (from Opinion 15/2011 Consent). Now what happens when the user (the "data subject" in terms of the Article 29 Working Party) does not consent to the Relying Party? The transaction gets cancelled, and the Relying Party will not be able to see any attributes (in legal terms: no personal data). That is an improvement on some current systems where the user has no direct control over his personal data and third parties might still be able to see those data after access has been denied. Regarding the addition from Article 7.4 of the pGPDR (on performing contracts and offering services even when the data subject refuses consent for the collection of irrelevant and unnecessary data), IRMA does currently not allow users to share some attributes while denying others: the Relying Party asks the user for (a list of) certain attributes, who can either agree to share them all and receive the service, or disagree and walk away. This is not in accordance with the mentioned condition in Article 7.4 of the pGPDR. See also the paragraph on user interaction (section 4.2.3.3). Again (as seen earlier with Mydex in section 4.1.4.1), the Relying Party can still force the user to enter their PIN by other (illegal) means, making it doubtful whether this requirement truly holds.

Specific "Blanket consent without determination of the exact purposes" is not valid. Applying this to IRMA, the specific purposes of the revelation of attributes must be known to the user when giving their consent to the Relying Party. As we have seen in section 4.2.2.1, all the Relying Party does is showing the cardholder which attributes it will check, who confirms this by entering their PIN to prove ownership of the IRMA card and to give their consent to the transaction. Taking a closer look at the *specific* requirement, the Working Party adds that the given consent "cannot apply to an open-ended set of processing activities. This means in other words that the context in which consent applies is limited." This seems true in IRMA: after entering a PIN, the attributes revealed are only processed within the transaction to which the user agreed. The Scheme Manager ensures (through certificates) that only the attributes that are asked for are being verified: when the IRMA card receives a request from the Relying Party, it first checks the certificate and extracts the public key and the attribute rights of this particular verifier. This ensures the IRMA card knows whether the requested attributes may be verified or not, and technically secures that only the data for which the individual gave consent are being verified. On the other hand it does not guarantee the storage of these attributes: it might be possible for the Relying Party to store attributes (longer than allowed) without asking the individual any consent for doing so. Storing these data might still lead to profiling, even without direct linkability to individuals. I will discuss the latter in the *purpose limitation*-paragraph (see 4.2.4.2).

Informed The issue we saw in the paragraph above applies here: the purposes of the processing are not fully disclosed to the user. This is in contradiction with the *informed* requirement stating that "the data controller [the Relying Party] must inform the data subject [the user] with the identity of the controller and the purposes of the processing, as well as any possible further information such as the recipients of the data." IRMA should guarantee this.

Other information such as the identity of the Relying Party and the recipients of the data are always known to the user since they will always use the IRMA

card directly with the Relying Party itself, and no third party other than the Relying Party will see the data. Should the Relying Party share these attributes outside the transaction would lead to revocation of its certificate provided by the Scheme Manager. The question however is how to detect this: as said before in the *specific*-paragraph, the Relying Party can always store attributes without knowledge of the individual or the Scheme Manager, and thus also share these with third parties (illegally, obviously).

Unambiguous As seen in section 3.1.1, unambiguous consent is defined by the Article 29 Working Party as consent "that leaves no doubt as to the individual's intention to provide consent." IRMA enforces unambiguous consent by requiring the individual to enter a PIN before any personal data are revealed.

Explicit IRMA also ensures that given consent is always *explicit*. By entering a PIN the user explicitly expresses "his/her wish to have his/her data processed for certain purposes." To explicitly know these "certain purposes" might require some form of interaction between the user and the Relying Party outside IRMA.

4.2.4.1.2 Contract A second legal ground is processing which is necessary for performance of, or entering into, a contract between data subject and data controller. An IRMA specific example would be a case where a user has signed a contract with a Relying Party, allowing the latter to process certain relevant data in exchange for a service or product. Consider for one example a contract where the Relying Party needs the address of the data subject to validate the contract: the user can load their address attribute onto their IRMA card and show it to the Relying Party.

4.2.4.1.3 Legal obligation As summarized in section 3.1, there are legal grounds other than plain consent or a binding contract for acquiring data, namely legal obligations, vital interests of the user, public interest in general and legitimate interests of the data controller. It should be noted that the need for a user to enter a PIN does not imply that these requirements do not hold: entering a PIN does *not* equal the legal meaning of consent[21], for an organization can demand the user to enter their PIN based on one of the other legal grounds. In those cases, entering a PIN can be seen more as providing access than as giving consent.

In order to obtain data based on a legal obligation, the Relying Party should be able to show these obligations to both the user and the Scheme Manager. Now we can actually look at this in the same way as at the contract requirement above: in combination with the contract everyone signs with IRMA in advance, this is compulsory for the user to enter their PIN - otherwise the servive is rejected. The aforementioned example of buying alcohol after showing the "older than 18" attribute is a typical example of data processing on the ground of legal obligation.

[21] Note that even in section 4.2.4.1.1 about consent, we *deduce* giving consent from entering a PIN.

4.2.4.1.4 Vital interests Vital interests relate to cases where the processing of data is necessary for the well-being of the data subject, another person or the public in general, in a situation where the data subject is not able to express their intentions. The latter seems a problem with IRMA: the user always has to enter their PIN to share attributes. In the aforementioned example (see also section 4.1.4.4 on Mydex) with an unconscious patient on the intensive care, there is no possibility for the doctor to acquire the blood type attribute from the IRMA card without activity of the user.

4.2.4.1.5 Public interest In cases where "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed" (Article 7 of the DPD), IRMA will still require the user to enter their PIN. This might not always be possible in cases of public interest where for example the police quickly needs personal data.

4.2.4.1.6 Legitimate interests In the case of processing personal data for the purposes of *legimitate interests* of the Relying Party, after a thorough balancing test has shown that those legitimate interests are not overridden by the interests and fundamental rights of the user, the user should always have the right to object against this decision. This is not a technical issue and thus lies not entirely within the scope of IRMA, but IRMA at least guarantees that:

- only the Relying Party can see the data,
- the user gets notified that the Relying Party processed their personal data, and
- no other data than relevant for the legitimate interests of the Relying Party will be shared with the Relying Party.

In short, those guarantees seem sufficient safeguards when processing data on the ground of legitimate interests of the data controller.

4.2.4.2 Purpose Limitation We have seen how IRMA deals with the legal grounds for data processing, but how does it deal with purpose limitation? In this paragraph I will take a closer look at the requirement (as stated in 3.2) that personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.

Specified According to the DPD, the data controller must "carefully consider what purpose or purposes the personal data will be used for, and must not collect personal data which are not necessary, adequate or relevant for the purpose or purposes which are intended to be served." Applying this to IRMA, this means it should not be possible for a Relying Party to use the attribute(s) it reads from an IRMA card for any purpose other than specified before the transaction. Credentials are issued by the Issuer using double-blind signatures, meaning the Issuer can no longer see the attributes within the credential after it

issues and signs one to an IRMA card[22]. This technically ensures that the Issuer is unable to trace (the use of) credentials and attributes. The Relying Party on the other side is able to decrypt the credential and to see the attribute(s) inside. (The Relying Party can obviously also verify the integrity of the credential through the digital signature, but that is less relevant within this context.) The IRMA architecture prevents the Relying Party from viewing this credential outside the transaction or storing it, thus technically not allows for using the credential for any other purposes. Of course unlawful use is always possible (for example: the Relying Party could install illegal, undetectable hardware which stores all attributes to profile customers), but requires extra effort and lies outside of the scope of IRMA. By default, IRMA handles the *specified* requirement well.

Explicit As seen in 3.2, the data controller must ensure that the purposes of data collection are "clearly revealed, explained or expressed in some intelligible form". This highly relates to what I discussed earlier at the *specific* requirement in 4.2.4.1.1: the Relying Party only shows the Individual which attributes it will check. Yet, the purposes might be easily derivable in most cases (for example: most people know it is a legal obligation to show they are over 18 years old when buying a beer, so the purpose of collecting the *over 18*-attribute is pretty straight-forward). There are also cases when one can not assume the purposes are explicitly stipulated, even when the purpose is obviously relevant or necessary (see *specified* requirement above) for processing. Consider for example revealing an address when placing an order: will the address be used solely for delivering purposes, or will the Individual also receive future junk mail from the Relying Party? The Relying Party should therefore always tell the Individual explicitly the purposes of collecting their data. By requiring the Individual to enter a PIN before revealing attributes, IRMA facilitates this as much as technically possible.

Legitimate The legitimacy requirement is the broadest of the three purpose limitation requirements, demanding that the purposes of the data collection are in accordance with all other laws. As we have seen in the paragraphs above, IRMA cannot guarantee that the Relying Party does not store attributes outside a transaction and thus use those for unlawful purposes. Within transactions the processing of data is lawful as long as the goal of the transaction is lawful: in order to make it possible to use IRMA, the Scheme Manager must have accepted the Relying Party in an earlier stage as a party that deserves a certificate. This makes it highly unlikeable that IRMA is being used for unlawful transactions.

Further processing Within the IRMA architecture, attributes can only be processed once, meaning that any issues with further processing lie out of the scope of IRMA. As said before, IRMA cannot guarantee that the Relying Party does not store attributes outside a transaction (hence outside the IRMA architecture) or processes them any further.

[22]Note that these attributes are still considered personal data: a Relying Party can view the attributes after decrypting them, and users can update/modify their attributes through the Attribute Issuer after authenticating (hence creating a link between the attributes and the data subject) themselves.

Feature	Supported by IRMA?
Data minimization	✓
End-to-end encryption	✓
Meaningful consent	
- Explicit consent	✓
- Specific consent	
Other grounds	
- Contract	✓
- Legal obligation	✓
- Vital interests	
- Public interests	
- Legitimate interests	
Purpose specification	
Purpose limitation	
Anonymization	
Pseudonymisation	✓
Legitimate profiling	
User can terminate data processing	✓

Table 2: IRMA round-up

4.2.5 Conclusion

A summary of the security and legal related features of IRMA can be found below in table 2. The key feature of IRMA seems to be data minimization, as we have seen that the data subject only shares (certain attributes of their) personal data that are relevant for the purposes of the data processing. Furthermore, we have seen that data are encrypted before being downloaded to the IRMA card, and only decrypted when shown to a Relying Party. The idea behind IRMA is that Relying Parties do not store these attributes, which, along with the possibility for the data subject to only share non-personal attributes, makes profiling hard. Yet, IRMA cannot guarantee full anonymization: for one thing because the data subject sometimes simply needs to share an identifier (citizen number, name, et cetera), but also since IRMA cannot fully guarantee different transactions cannot be linked *outside* of the IRMA environment. Other problems that remain are the lack of purpose specification, purpose limitation and dealing with some other legal grounds for data processing.

4.3 Synergetics

Synergetics is a fourth Personal Data Management system, much like the other systems based on the concept to give (back) the individual their own control over their own data. Their so-called *end2end Trust Assurance framework* (from now on referred to as *end2end*) is a commercial implementation of the *Trusted Architecture for Securely Shared Services*[23] project (from now on referred to as *TAS*^{3}), which ran from 2008 through 2011.

One running example of Synergetics focuses on the health care sector, where the main idea is to make patients themselves the center of their own - often highly sensitive - medical data, giving them full control over what they share with whom. All data are also anonymized for research goals to help medical professionals and to maximize effective treatments. In this chapter I will take a closer look at the end2end framework and the specific health care example, also known as the *ReLifE Health Care Ecosystem*. The following information is all derived from [Kellomäki and Vervenne, 2013] and the Synergetics website[24].

4.3.1 Business case

Although started as a research project by (among others) TAS^3, several universities and Synergetics, the end2end system can be seen a commercial implementation of a PDM. This is partially shown by Synergetics offering the service to all kinds of networks, mainly focusing on health care and business communites. But it particularly shows in the design of the end2end environment itself: different than other PDMs, Synergetics derives anonymized (which will be discussed later) data from the personal data stores themselves, and offers these data to third parties.

4.3.2 Actors

4.3.2.1 Users In the health care example, the **data subjects** ("users" within the framework of Synergetics) are generally patients and other people with medical records. They use the ReLifE Health Care Ecosystem to conveniently share their health data like (change in) weight and self-measured blood values. To ensure that patients have control over their own medical data, they decide what data they share with whom through their Personal Data Store (PDS). An interesting feature of the end2end environment is the possibility for users to delegate responsibilities to another user. This is especially useful in health care examples where the user can no longer handle their own (privacy-related) issues concerning data sharing, but can still appoint a trusted person to handle these for them.

4.3.2.2 Service providers Again focusing on the health care example, service providers can be roughly divided into two groups: practitioners (doctors, nurses, other hospital and pharmacy staff) who need patient data in order to treat them, and "other" third parties (for example insurance companies) who are mainly interested in the patient data for their own good. All service providers sign contracts with the end2end environment that oblige them to keep

[23] TAS^3 - `http://www.tas3.eu/`
[24] Synergetics BV - `http://synergetics.be/`

to agreements, even when those agreements are beyond the compliance of the technical framework. As in other PDMs, all service providers can be seen as **data controllers**.

4.3.2.3 end2end governance The governance and administration of the policies in the trust network occur on layers on top of the service providers and (end) users, as is shown in figure 5. All the actors involved sign contracts with each other: I have not been able to view these contracts myself, meaning I will assume in this chapter that these contracts are all legally binding and ensure accountability in case one of the actors (service providers, (end) users or PDM) violates the relevant terms and conditions.

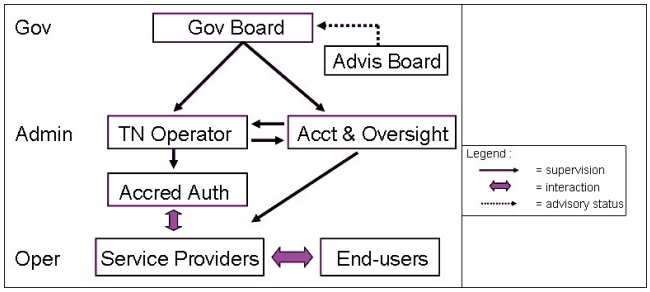

Figure 5: General overview of the trust network within end2end environments (source: [Synergetics, 2014])

4.3.3 Framework

In this paragraph I will take a closer look at the framework of the end2end environment of Synergetics. I will do so by using the approach as suggested by [Bus and Nguyen, 2013] and divide the end2end environment into infrastructure, data management and user interaction as introduced in chapter 2.1.

4.3.3.1 Infrastructure All personal data are encrypted and stored within the secure Personal Data Store of the user, and remain encrypted during data transaction with service providers. Taking a closer look at these data transactions, we see that Synergetics uses *Sticky Policies*: data sharing terms, agreements and policies are *stuck* on top of the data. These data then can only be decrypted by a third party when they accept these terms. According to [Kellomäki and Vervenne, 2013], these policies are stored somewhere in the "audit summary" and seen as a binding contract between user and service provider.

Access to profile information and the personal data store requires two-factor authentication using Yubikey Single-Sign-On[25]. Single-Sign-On means that only one authentication event is required to access multiple services or applications. To guarantee the security of this Single-Sign-On system, both a physical token and a credential are required to authenticate (hence *two-factor* authentication).

The end2end environment offers a built-in audit trail that explicitly shows the entire communication history between user and service providers. With this

[25]Yubikey - `http://www.yubico.com/applications/single-sign-on`

audit trail, the end2end system goes beyond the measures we have seen so far: an online conformity test will continuously automatically check whether service providers comply with their commitments. Those commitments are defined beforehand in a contract between Synergetics and the service provider. The check is performed by creating fake users whose data are being tracked throughout the system: an example infringement would be data that are stored longer than the sticky policy (see also section 4.3.3.2) allows for. Identified infringements are forwarded to an accountability committee which is part of the ecosystem governance structure, in order to initiate corrective actions, inform the concerned users and, if necessary, sanction the wrongdoer. At the moment of writing this online conformation test is still under construction, and no technical details are known.

4.3.3.2 Data Management Personal data are stored in multiple online, cloud-based Personal Data Stores. Users can pick the PDS of their choice and can decide to change from one to another at any given time. Every website or service provider has to ask permission to pull data from these stores, which is described by [Kellomäki and Vervenne, 2013] as the *pull model*, as opposed to a *push model* where a user would offer their data to a third party.

Synergetics distinguish themselves from other PDMs by performing analytics on the ((pseudo-)anonymised) user data, and then offer these analytics to the companies and service providers. At the time of writing it is still being researched how the anonymization process can be made entirely safe. Should they succeed, this analysis feature would be a big step in convincing governments and organizations to use PDMs.

4.3.3.3 User Interaction The user connects to a service provider which typically offers a service in return for sharing personal data. When asked for personal data, the user authenticates through the identity provider of their choice. After the user has picked their IDP, they are sent to the corresponding login screen. The login screen of the IDP is the only place in the system where a user enters their credentials. Within the end2end framework, credentials can vary from hardware tokens and electronic identity cards to straightforward username and password authentication. After succesful authentication the user gets redirected to the service provider.

Apart from interaction with the current service provider, the user can decide to use one of the web services from the ReLifE system or consult their own *Personal Trust Manager* (from now on referred to as *PTM*). This PTM stores all former web service connections and data sharing permissions, and lets the user manage, add and/or withdraw those permissions through a frame within the website. The personal data themselves are stored elsewhere in the *Personal Data Store* (or *PDS*) of the user. The user always has to perform an action in order to transfer data from this store to a service provider.

4.3.3.4 Data types When we look at the different types of data that are possibly involved in data processing, Synergetics enables the processing of volunteered data like age, name, address and health data, as well as observed data like web browsing. Furthermore, Synergetics also offers inferred data (statistics and other data analysis) to third companies. This makes Synergetics the

only (discussed) PDM that handles all three types of data. Note however that Synergetics are inferring these data themselves, instead of dealing with inferred personal data. A PDM that deals with raw metadata is discussed in section 4.4.

Contrary to the other PDMs discussed here, the ReLifE Health Care Ecosystem focuses on health data. This leads to a different approach for checking the legal compliance of Synergetics, as can be seen in section 4.3.4.1 below.

4.3.4 Compatibility with Legal Requirements

Within the ReLifE Health Care Ecosystem it seems obvious that the end2end environment deals with personal health data. Yet before taking a look at the compatibility with the legal requirements, I will first look at the DPD and Article 29 Working Party to see what they consider to be health data, and what not. I will then continue discussing the legal requirements as I did with Mydex and IRMA.

4.3.4.1 Health data The Article 29 Working Party considers[26] personal data health data in the following cases:

1. The data are inherently/clearly medical data
2. The data are raw sensor data that can be used in itself or in combination with other data to draw a conclusion about the actual health status or health risk of a person
3. Conclusions are drawn about a person's health status or health risk (irrespective of whether these conclusions are accurate or inaccurate, legitimate or illegitimate, or otherwise adequate or inadequate)

Since all three cases hold within the ReLifE environment, we can safely assume that Synergetics deals with health data, and a more detailed look at the DPD is needed.

As summarized in section 3, Article 8(1) of the DPD states that "Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, *and the processing of data concerning health* or sex life", unless at least one of the five exceptions named in Paragraph 2 of the same Article hold:

1. The data subject has given their explicit consent;
2. Processing is necessary for carrying out a legal obligation;
3. Processing is necessary for carrying out vital interests;
4. Processing is necessary for legitimate interests;
5. Processing is related to data that are manifestly made public by the data subject or processing is necessary for the exercise of a legal claim.

[26] 05/02/2015 Letter from the ART 29 WP to the European Commission, DG CONNECT on mHealth - `http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf`

For a more detailed description of terms like explicit consent, legal obligation and legitimate interests, I refer to section 3.1. I will check whether Synergetics complies with these legal grounds for the processing of health data in the following paragraphs, but first need to address one important exception to these grounds. As Article 8(3) states, none of these legal grounds is needed as long as "processing of the data is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services". This is relevant to the Synergetics PDM, since all processed data are *intended* for medical diagnosis and treatment by medical professionals. Note however that all data are also being processed by Synergetics themselves, and therefore still need to comply with at least one of the legal grounds from Article 8(2) mentioned above.

4.3.4.2 Explicit consent Quoting the DPD, "an active response, oral or in writing, whereby the individual expresses his/her wish to have his/her data processed for certain purposes" is required, whereas a pre-ticked box can never lead to explicit consent being given, as it does not involve some positive action from the data subject (see section 3.1). Looking at Synergetics, we saw (in section 4.3.3.3) that the user always has to perform an action to transfer data from their data store to a service provider. It however remains doubtful whether this action can be seen as "an active response" as required by the DPD: within the environment of the personal data store, a user can share their (sensitive, medical) data with a few clicks.

4.3.4.3 Legal obligation Synergetics has the same way of dealing with the ground of legal obligations as Mydex (see 4.1.4.3): a legislative body or government can oblige the individual to enter their credentials and agree to share their data with the data controller, but Synergetics has no feature to automatically facilitate this. Note that releasing data in this context is significantly different than giving consent.

4.3.4.4 Vital interests Other than Mydex, Synergetics provides a way of sharing data when "the data subject is physically or legally incapable of giving his consent" (Article 8(2)(c) of the DPD): as mentioned in 4.3.3.3, the user can (when psysically capable, obviously) appoint a trusted person beforehand who can transfer data from the PDS to the service provider. Apart from sharing, the trusted person is not able to upload data to the PDS or modify anything themselves, ensuring the integrity of the data. This seems to be a good solution in handling with the legal ground of vital interests.

4.3.4.5 Legitimate interests Synergetics does not seem to include the possibility to balance the legitimate interests of a third party against the interests and fundamental rights of the data subject. This balancing test should now be done outside the scope of Synergetics, after which the user can be obliged to share their data by transfering data from their data store to the service provider.

4.3.4.6 Pseudonymisation While all PDMs are in fact data controllers, Synergetics goes one stup further and processes the data themselves by offering anonymized personal data to scientific and medical researchers. As mentioned

in section 3.4, most data protection principles of the DPD and pGDPR do not apply when it comes to fully anonymous data. It is therefore extremely interesting to see what Synergetics exactly does with the personal data. Unfortunately, at the moment of writing there is no technical implementation of the anonymisation process yet. I will therefore distinguish two possible outcomes for the anonymization process. Case one is where all personal data are fully anonymized, meaning no research data can be linked to any specific person, and no further compatibility with EU legislation is required. In the more likely case of the data only being *pseudonymised*, the data protection requirements regarding profiling apply.

4.3.4.7 Purpose limitation One of the key features of the end2end environment is the use of sticky policies. As we have seen in section 4.3.3.1, these policies are pinned on top of the data, and third parties have to agree to these policies in order to decrypt the data. Unfortunately, this is not further nor technically specified. If we assume that all purposes (both in goals and duration) for the collection of personal data are specified within these sticky policies, and that these sticky policies prohibit the decryption of the payload beyond this specified point in time, the requirement of purpose limitation can be met. Should a third party store the decrypted data outside of the environment for longer than allowed, or use these data for other purposes than specified, the contract between third party and Synergetics would be violated and the third party would be accountable. Although the end2end audit trail (supposedly) is able to detect such fraud *within* the environment, it seems unlikely that data stored *outside* of the PDM are still traceable.

4.3.5 Conclusion

A summary of the features of Synergetics can be found below in table 3. In the previous sections, we have seen a couple of interesting things. One of them is the - yet to be implemented - automated online conformation test to check for "infringements" and whether third parties comply with their commitments. Although still in the implementation phase, this could ultimately lead to a system that deals well with the relevant purpose specification and profiling requirements from the DPD and pGDPR. A second interesting feature is the "anonymization" algorithm, which Synergetics applies to the personal data themselves. As I discussed earlier, we should see this more as pseudonymisation since all "anonymized" data can be linked to a data subject once one or more identifiers are used. Regarding the other requirements, it remains doubtful whether the implementation of "giving consent" can be seen as an active response from the user as required by the DPD: within the environment of the personal data store, a user can share their (sensitive, medical) data with a few clicks. Vital interests on the other hand can be seen as a succesfully implemented legal ground through the trusted person feature. Sticky policies are used to ensure purpose limitation, which at least (if technically implemented correctly) guarantees that third parties do not store or process data within the environment for goals or time periods other than specified. Once again, the golden rule ("once the data are out, they are out") applies: data can still be decrypted for valid purposes, and then stored or processed outside the end2end environment for other means.

Feature	Supported by Synergetics?
Data minimization	✓
End-to-end encryption	✓
Legal grounds for processing	
- Explicit consent	
- Legal obligation	✓
- Vital interests	✓
- Legitimate interests	
Purpose specification	
Purpose limitation	
Anonymization	
Pseudonymisation	✓
Legitimate profiling	
User can terminate data processing	

Table 3: Synergetics round-up

4.4 openPDS/SafeAnswers

We have now seen several types of PDMs: centralized data environments, attribute-based PDMs on smart cards, health care environments which infer data from anonymized user data, etc. Yet we have not seen a PDM that specifically focuses on inferred personal data. There is one: a combination of so-called *openPDS* and *SafeAnswers* purely concentrates on personal metadata. Similar (yet different, as shown later) to Mydex, openPDS can be seen as a trust framework that allows users to store, control and share their own metadata. SafeAnwers is a technique to transform these possibly high-sensitive, easily identifyable metadata into "low-dimensional answers that are less likely to be re-identifiable and to contain sensitive information" ([de Montjoye et al., 2014]). These *answers* can then be integrated into the same openPDS environment and easily shared with third parties like mobile applications and web browsers.

4.4.1 Business case

openPDS/SafeAnswers is intellectual property of the Massachusetts Institute of Technology[27] and distributed under a Creative Commons license. The latter means that all current research documents and software are openly available for public access and (re-)use. This is in line with the vision of the developers as described on the website of openPDS/SafeAnswers[28], stating the following:

> We believe that a New Deal on data is needed. When it comes from data, "ownership" should to be thought of according to the old English common law. Data ownership would therefore be defined as the rights of possession, use, and disposal instead of a literal ownership. Discussions on such changes and their implications for privacy must

[27] Trust Framework System Rules for Personal Data and Individual Identity Services - `http://openpds.media.mit.edu/documents/System_Rules.pdf`

[28] openPDS/SA - `http://openpds.media.mit.edu/`

> also take into account the current political and legal context. (...) [This context recognizes] the increasing need for personal data to be under the control of the individual as he is the one who can best mitigate associated risks.

4.4.2 Actors

4.4.2.1 Users Without openPDS/SafeAnswers, users share their raw metadata directly with third parties such as mobile applications. An example is given by [de Montjoye et al., 2014] where user Alice wants to install the location-based check-in application LBSinc for her smartphone. In the old model (without openPDS), "Alice downloads the app onto her phone, authorizes LBSinc to access her phone's network communication and GPS coordinates, and creates a user account with LBSinc. The LBSinc app starts collecting metadata about her and stores it all in its back-end servers. Under this model it is difficult for Alice to access the metadata LBSinc uses to makes inferences about her, or to remove the metadata she does not want LBSinc to access or use." Using openPDS, users share these raw metadata with the PDM instead. In the LBSinc-example, Alice downloads and installs a PDS-aware version of LBSinc, and authorizes it to access only her phone's network communication (as opposed to also sharing her GPS coordinates). When using the app for the first time, LBSinc will ask Alice to enter the location of ther openPDS-account, allowing her to see exactly what (parts of) metadata will be used. Based on this information, she can then decide to either accept or deny using the app.

4.4.2.2 App providers Within openPDS/SafeAnswers, applications will no longer be able to collect all the raw user metadata they desire. Instead, they use a SafeAnswers module to *ask* the metadata for relevant, anonymized information. This actually offers several advantages to app developers and providers, as is shown by [de Montjoye et al., 2014]: by using openPDS/SafeAnswers, a lot of metadata have probably already been collected by the PDM or other applications. This implies that the developer has access to a potentially large metadataset, including historical data, saving them time on deducing relevant information, anonymizing metadata, securely storing and starting from zero themselves. Instead, the app developer has to write a SafeAnswers module to extract the relevant metadata from the PDS.

4.4.2.3 Researchers As seen with Synergetics, scientific researchers are possibly very interested in metadata. Just like appplication developers, researchers can use openPDS/SafeAnswers by writing SafeAnswers modules to get relevant, anonymized information. In a test case described by [de Montjoye et al., 2014], a first field study has been held monitoring the daily (smartphone) behavior of people diagnosed with mental problems, where anonymized metadata are observed "to reproduce the diagnoses of mental health conditions, focusing on changes in speech and social behavior".

4.4.3 Framework

In this paragraph I will take a closer look at the technical framework of openPDS/SafeAnswers. I will do so by using the approach as suggested by [Bus

and Nguyen, 2013] and divide the openPDS/SafeAnswers framework into infrastructure, data management and user interaction as introduced in chapter 2.1.

4.4.3.1 Infrastructure The openPDS environment consists of two main parts: the front-end and the database. The front-end functions as the connection between raw metadata in the database and interested parties outside the openPDS. In the front-end, SafeAnswers modules are executed (within a separate safe, closed environment) on the metadata they are allowed to read. The SafeAnswers (results of these modules) then leave the PDS through the front-end, where all communication is encrypted using 256 SSL connections. This all implies that third parties never have direct access to raw metadata in the database, but only communicate through SafeAnswers modules. This is also illustrated by figure 6 (source: [de Montjoye et al., 2014]).

An example of this architecture could be a smartphone app which gives a user specific diner recommendations based on their location. Each time this application needs to recommend a restaurant, it sends a request to the openPDS of the user. The SafeAnswers module of the application inside the front-end then accesses the required metadata, computes the answer (in this case the name and location of a restaurant) within a safe environment and sends this back to the application. The application itself never has access to any metadata in the database.

4.4.3.2 Data Management Within the database, raw metadata are stored in a CouchDB database. Apache CouchDB databases are so-called NoSQL (or *Not Only SQL*[29]) databases, where data and relations are - unlike in common relational databases - stored as collections of independent documents. Besides improved speed and availability, the main advantage of CouchDB is the large range of existing functionality to support the use of SafeAnswers modules, including support for MapReduce functions and data validation.

4.4.3.3 User Interaction As summarized in section 2.1, [Bus and Nguyen, 2013] suggest that each personal data management system should offer simple and intuitive tools for the user "to have meaningful interaction with service providers regarding the permissions and policies associated with the use of their personal data." Within openPDS/SafeAnswers, the user has this kind of meaningful interaction before granting access to an application. See also figure 7 (source: [de Montjoye et al., 2014]): the PDS-aware version of the application shows what types of questions it will ask from the metadata of the user, as well as example responses and the sensors used to compute these responses.

4.4.3.4 Data types This is the key aspect of openPDS/SafeAnswers: it focuses on inferred data. Instead of the more common approaches we have seen so far, there are no volunteered data involved. Observed data are shared directly with the openPDS, where SafeAnswers modules connect third parties with these data by inferring only relevant and allowed data from the database. This is different than Synergetics, where the PDM itself infers data. Within

[29]Definition of NoSQL - `http://searchdatamanagement.techtarget.com/definition/NoSQL-Not-Only-SQL`

openPDS/SafeAnswers, it is still the applications and third parties that write the modules and infer the data, yet now within the secure and controllable environment of a PDM.

4.4.4 Compatibility with Legal Requirements

Since openPDS/SafeAnswers focuses on observed and inferred data, it is particularly interesting to see whether this PDM adds something on the legal requirements regarding profiling. As described in section 3.4, profiling based solely on the processing of pseudonymous data may not presumed to be significantly affecting the interests, rights or freedoms of the data subject. The question then remains whether the data inferred by openPDS/SafeAnswers are truly pseudonymous.

4.4.4.1 Pseudonymity The main idea is that within openPDS/SafeAnswers only *answers*, not *data*, are collected. Or, as openPDS/SafeAnswers state on their website[30]: "Rather than exporting raw accelerometer or GPS data, it could be sufficient for an app to know if you're active or which general geographic zone you are currently in. Instead of sending raw accelerometers readings or GPS coordinates to the app owner's server to process, that computation can be done inside the user's PDS by the corresponding Q&A module." That looks promising when we consider the aforementioned study by [de Montjoye et al., 2013] (yes, that is the same person as the one behind openPDS/SafeAnswers) which concluded "four spatio-temporal points are enough to uniquely identify 95% of the individuals" (see also section 3.4). One could however argue that these "SafeAnswers" are personal data nonetheless, which are out of sight and control once they leave the personal data store of openPDS.

4.4.4.2 Suitable measures

Notification The first requirement when it comes to automated processing is, that the data subject will always obtain "confirmation as to whether or not data relating to them are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed" (Art. 12 of the DPD). As we have seen in section 4.4.3.3 and in figure 7, the data subject (or "user" in terms of openPDS/SafeAnswers) always gets confirmation when data are being processed, as well as a notification of the categories of data concerned (for example GPS data, phone call history or social media behavior) and the recipients of the data (for example Foursquare or a research app). The purposes of the processing seem, although often easily derived from the goal of using the application, less specifically defined.

Intelligible form Secondly, Article 12 of the DPD demands that the data subject always receives communication "in an intelligible form of the data and of any available information as to their source". As described in section 4.4.3.3 about user interaction, openPDS/SafeAnswers presents all relevant communication to the data subject in an intelligible form. In an easily understandable

[30] openPDS/SafeAnswers - `http://openpds.media.mit.edu`

environment, users can see what kinds of metadata are observed in order to answer questions to SafeAnswers modules.

Knowledge of the logic involved Especially interesting when processing observed and inferred data is providing the data subject with "knowledge of the logic involved (...), presented to the data subject in an intelligible and understandable form" (Art. 12 of the DPD). Within openPDS/SafeAnswers, this "logic involved" equals SafeAnswers modules. These modules themselves are programmed scripts which operate on Design Documents in the CouchDB database, hence not something the regular data subject would understand. As mentioned earlier, these SafeAnswers modules are presented in a more understandable way such as in figure 7. Yet the logic itself is currently not included. Considering the example, it would be pertinent if the application also showed *how* the exact "how social are you?" score is being calculated from the call, SMS and bluetooth logs of the data subject.

4.4.4.3 Purpose limitation

Specified The purposes for collecting the data must always be specified clearly. As said in section 4.4.4.2, users always get confirmation of both the data concerned and the recipients of the data. The first must ensure that data controllers will not collect data that are unnecessary for the purpose of the data processing: there is no need for the aforementioned example check-in application to access call histories or photo storage on the device of the data subject. The DPD also requires the use of adequate data: a requirement hard to completely ensure, but openPDS at least shows the data subject the time periods for each category of data it collects. This does not entirely solve the *adequate* issue, but at least allows the data subject to deny data sharing if they feel the release of inadequate data is being asked.

Explicit The purposes for data collection are clearly revealed and expressed in some intelligible form, as we have seen in section 4.4.4.2.

4.4.5 Conclusion

A summary of the security and legal features of openPDS/SafeAnswers can be found below in table 4. As mentioned in the purpose limitation paragraph (section 4.4.4.3), the data subject can terminate data sharing at any given time. openPDS/SafeAnswers also shows the user which attributes are being shared and for how long, although the exact purposes should still be specified further in order to comply with the relevant DPD requirements about purpose specification and meaningful consent (note that the latter requires specific consent for specified purposes). A bigger problem with openPDS/SafeAnswers is that it aggregates a lot of data about one user, although perhaps anonymised at the beginning, that later can easily be used to profile the data subject. This allows for specific targeting, without the possibility for the user to get detailed insight in or control of their own profile, let alone the possibility to monetize their own personal data.

Feature	Supported by openPDS/SafeAnswers?
Data minimization	✓
End-to-end encryption	✓
Meaningful consent	✓
- Explicit consent	✓
- Specific consent	✓
Insight in automated processing	
- Notification	✓
- Intelligible form	✓
- Knowledge of logic involved	
- User can rectify data	
- User can erase or block data	
Purpose specification	
Purpose limitation	
Anonymization	
Pseudonymisation	✓
Legitimate profiling	✓
User can terminate data processing	✓

Table 4: openPDS/SafeAnswers round-up

5 Summary and conclusion

In the previous chapter, I have discussed four PDMs and compared them on their features and their compliance with legal requirements. Before diving into conclusions and answers to my research questions, in this section I will summarize how these PDMs comply with these legal requirements. The result is a first, general overview of which legal requirements can easily be met by PDMs, and which seem harder to fulfill, need further policies or cannot be met at all. This will lead to a more specific overview of features that are solved, *not* solved and created by using PDMs, which will answer the research questions I asked in the beginning of this thesis.

5.1 Legal compliance

5.1.1 Consent

One of the legal grounds for data processing is consent. As we have seen, most PDMs work from an environment in which users share their own attributes with whomever they prefer. We have seen different approaches: PDMs like Synergetics and openPDS/SafeAnswers ensure that given consent is always unambigious and explicit by requiring users to tick a box, click "I agree" or do something similar after showing the data subject the corresponding processing purposes. Whether that is also the case for the "freely given" requirement is another question. Within the scope of the PDM, users will always decide themselves whether they give consent or not. That is not all: freely given also means that refusing consent to the processing of unnecessary data is not a valid ground for the data

controller to abort (further) provision of a service or execution of a contract. We have seen how Mydex handles this, by allowing data subjects to allow or deny permissions one by one, even within one data connection. The Mydex connection between the data subject and a controller then persists even if the former should decide to deny all possible data sharing permissions, in accordance with the requirement. This could be an imporant PDM feature in dealing with mobile apps that require all sorts of unnecessary (for example GPS and address) data. Yet, as I have mentioned earlier in the discussion of the single PDM examples, users can still be forced to give consent by other, illegal means.

In a second type of PDM, IRMA in this example, we have seen how the data subject carries their attributes (offline) with them on a smart card, and only shares specific attributes during a transaction with the data controller. By requiring the data subject to enter their PIN before the transaction, IRMA ensures that consent is always unambiguous and explicit. We have seen that IRMA scores less well on the "freely given" requirement, since it does not specifically allow users to share attributes one by one. In the specific example of IRMA, a Relying Party asks the user for (a list of) certain attributes, who can either agree and receive the service, or disagree and walk away. This should not necessarily be standard procedure in attribute-based PDMs on smart cards, but such PDMs seem less useful in the aforementioned example of dealing with mobile apps. Since these *smart card* based PDMs often require some kind of face-to-face communication, the specific and informed requirements are not necessarily technically enforced by the PDM itself: in such cases, the data controller (or someone working on behalf of the data controller, for example a cashier) should fulfill these requirements by informing the data subject of the exact purposes of the processing before the data subject gives their consent.

To summarize: most PDMs offer good solutions for handling the legal requirement of consent, although they can still improve by technically enforcing that consent is always specific and informed.

5.1.2 Contract

The second legal requirement I discussed is based on contracts, or data processing which "is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract" (Article 7 of the DPD). PDMs can facilitate this in several ways. One of them is where the PDM itself is able to share user data with the data controller if a valid contract between subject and controller is presented. One obvious example is the Data Sharing Agreement in Mydex (remember that PDMs are data controllers themselves as well), although the actual data sharing still requires several clicks from the user. Both the data subject and data controller sign separate contracts with the PDM themselves, which gives the PDM the right to participate as a middle man in processing data in these cases. The key point here is that the PDM cannot view the actual, encrypted data themselves.

Even when - like in Mydex - the data subject is still required to release their data themselves, we have seen that releasing data is not the same as giving consent (see 4.1.4.3 and 4.2.4.1.3 for examples). The contract obliges the user to enter their PIN (in IRMA for example) or allow data sharing (in Mydex), after which the PDM basically (and technically) continues in the same way as if

the data subject gave their consent for data sharing (but for different grounds).

One further advantage of a PDM in such cases is that the data subject will always be notified when their data are being processed, unlike in a PDM-free world where the subject often has no idea when data are being processed based on contract. Additionally, PDMs could also implement a feature where data subjects can directly file a complaint in cases in which they feel data processing has been unlawfully based on the contract ground.

5.1.3 Legal obligation

Other legal grounds for data processing can be dealt with in similar ways as described above. When shown a valid legal obligation, a PDM could either share personal data of the data subject themselves (again, without seeing the actual data) or oblige the data subject to release the demanded data.

A common example of data processing on the ground of a legal obligation is buying alcohol, when the buyer is required to verify they are over 18. We have seen that IRMA uses this as their primary example: the user downloads the "over 18"-attribute onto their IRMA card, and shows the attribute to a Relying Party when buying alcohol.

Apart from such clear *user-shares-their-own-data* examples, it would be interesting to see how PDMs (can possibly) deal with more complex legal issues. Take for example tax return data: an ideal PDM shows which tax official is looking at your file, which *attributes* (to stick with IRMA terminology) determine the outcome of the tax return, et cetera.

5.1.4 Vital interests

As we have seen in section 3.1.4, the Article 29 Working Party defines vital interests as "essential individual interests of the data subject or of another person and it must in the medical context be necessary for a life-saving treatment in a situation where the data subject is not able to express his intentions." As mentioned earlier, we have not yet seen an implementation of a PDM where data can be accessed without some kind of action from the data subject. Again, the solution lies in the possibility for the PDM to share personal data with a data controller. Access to this feature could then be restricted to (for example) surgeons, ambulance and intensive care personnel only, to lower the risk of unwanted use access by third parties. It is therefore necessary to always notify both the PDM and the data subject of any data processing based on this ground.

In smart card based PDMs like IRMA, this legal requirement is impossible to comply with, since the data subject is always required to enter their PIN.

5.1.5 Public interest

Within the scope of PDMs, processing based on the legal ground of public interest is very similar to that based on a legal obligation.

5.1.6 Legimitate interests

One of the most interesting legal grounds for data processing is described in Article 7(f) of the DPD, and known as data processing on the ground of legitimate interests of the data controller. As we have seen in section 3.1.6, this

legal ground requires a balancing test between the legitimate interests of the data controller and the interests of the data subject: the data controller must always specify the purpose of processing, while the data subject is not required to specify why they would not wish to have their data processed.

In the PDMs I discussed earlier, I have not seen a full implementation of such balancing tests. Right now, there is no PDM which lets a data controller suggest their legitimate interests for the processing of user data.

5.1.7 Purpose limitation

As we have seen, none of the discussed PDMs complies with all the legal requirements concerning purpose limitation. Yet, they offer some solutions: as for specifying the purposes of the data processing, Mydex obliges third parties to specify these purposes through the Data Sharing Agreement, openPDS/SafeAnswers informs the user with both the data concerned and the recipients of the data, and Synergetics uses sticky policies containing these purposes on top of the encrypted data.

Such data sharing agreements and sticky policies must ensure that data stay adequate and accurate and that data are only being processed for the specified purposes, but do not solve the biggest purpose limitation issue: even an automated audit trail (like in Synergetics) which detects fraud will not be able to trace such fraud *outside* the environment of the PDM. In other words: even if the PDM ensures that all data are only being processed for valid purposes, it cannot prevent the same data from being stored and/or processed outside the PDM.

5.1.8 Automated processing

When it comes to decision making based on the automated processing of data, additional requirements are defined in Article 12 and 15 of the DPD (see also section 3.3). We have seen that legal grounds for decision making based on automated processing involve consent, contract and legal obligations: legal grounds for data processing that have already been discussed in the sections above. But we have also seen that when it comes to decision making based on automated processing, additional suitable measures have to be taken:

- confirmation as to whether or not data relating to them are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed,
- communication to them in an intelligible form of the data and of any available information as to their source,
- knowledge of the logic involved in any case of automatic processing of data. This logic should also be presented to the data subject in an intelligible and understandable form, and is especially interesting when it comes to observed and inferred data.

Especially with openPDS/SafeAnswers, we have seen how a PDM can provide a solution in dealing with these additional requirements: openPDS/SafeAnswers

communicates all data processing purposes, categories and recipients to the user in an intelligible form (through a mobile application, see 4.4.4.2), as well as (some of) the logic involved in the SafeAnswers-modules which extract the "answers" from the (observed and inferred) personal data. Although openPDS/SafeAnswers has not yet tackled all the problems (for instance: how are the exact scores being inferred from the observed data?), it has shown that a PDM can give the data subject a lot more insight in the processing of observed and inferred data.

5.1.9 Pseudonymisation

Some of the discussed PDMs already implemented pseudonymisation. Synergetics for example, as is described in section 4.3.4.6, infers pseudonymous data from the personal data stores of their users and offers the results to scientific (health related) research. Keeping data really anonymous within the scope of the PDM is not something that the discussed PDMs seem to be able to. We have seen for example how all data in IRMA only remain anonymous until an identifier (be it a citizen number, telephone number, name, et cetera) has to be shown, and how the SafeAnswers themselves can be seen as personal data, even when they are meant to be anonymous and unlinkable to the data subject. Note that this is not necessarily bad: even with PDMs, scenarios will exist in which identification of the data subject is desirable. At least most of the discussed PDMs offer pseudonymity to some extent, which is better than none.

5.2 Data types

5.2.1 Volunteered data

The majority of the Personal Data Management systems I discussed shares the same principle: the user chooses which data they want to share with whom until when, and the PDM then technically ensures that third parties only see the data what and when they are allowed to. Examples range from basic info like name, age and address to any manually entered data like energy use, pictures or billing information. As we have seen in section 1.1, these kind of data are basically known as *volunteered* data. Combining this with the discussed legal framework on data protection (see section 3.1), dealing with volunteered data is especially (but not only) related to two of the six legal grounds for data processing: consent and contracts. As we have seen, all PDMs score reasonably well on handling these legal grounds. Yet, some underlying consent-requirements like *specific* and *informed* tend to lag behind in those ratings. As discussed in section 3.1.1, those require that data controllers inform the data subject with the identity of the controller and the scope, consequences and purposes of the processing, as well as any possible further information such as the recipients of the data. Consent may never apply to an open-ended set of processing activities. The ideal PDM would implement this requirement and ensure that volunteered data are handled in perfect accordance with existing data protection legislation.

5.2.2 Observed data

What can PDMs do for *observed* data? As we have seen in section 1.1, this type of data is defined as data captured by recording activities of users, such as

cell phone location data, search histories, real time measured energy usage and digital cookies. Some PDMs already facilitate the processing of observed data. For example, Mydex features the possibility to automatically share real-time data (like energy usage, credit card transactions and phone call data) with user-selected third parties. Moreover, note that Mydex is not able to view the data themselves: data are uploaded, stored and shared cryptographically secure, and only viewable for the third party that is allowed to. In openPDS/SafeAnswers, we have seen how observed data are shared directly with the Personal Data Store. The observed data are then being pseudonymized by using SafeAnswers modules to ensure that only relevant (adequate, accurate, et cetera) data are inferred (see also next paragraph) and shared with connected third parties.

These examples show that a PDM can function as an important mediator between observed data and the data controller. Integrating this into the environment of an PDM has several advantages:

- Users will always be notified when observed data are being processed and by whom, hence improving transparency. Compare this to a PDM-less situation when users have zero to little insight in what data are being observed and by whom they are being processed.
- Users will notice and be able to take action when data are being processed unlawfully.
- Organizations, governments and other third parties benefit as well. Improving the transparancy and security also means an improvement of the integrity and accuracy of the data.

5.2.3 Inferred data

As I said earlier in section 1.1, inferred data are the result of "analysis of personal data" using data mining technologies on (collections of) volunteered and/or observed data. Examples are future consumption prediction techniques or credit scores. Without PDMs, these are the hardest data to control, as users often have no insight in when, what, how and by whom inferred data are being processed. PDMs can help when all behavioral patterns are *inside* the data management tool. Take for example Mydex, where (as we have seen with observed data) energy use can be uploaded directly to the data store. Hypothetically, interested parties can then infer behavioral data from these data within the data store, automatically notifying users and - if necessary - asking them for permission.

We have seen two PDMs that "do something" with inferred data. Synergetics infer data themselves but do not collect, process or control inferred data. openPDS/SafeAnswers can be seen as the only Personal Data Management system that currently tries to deal with inferred data, and does so quite well: we have seen how they pseudonymise the data as much as possible for the intended purpose of the data collection, and how they rate well measured against the legal requirements related to automated processing. A PDM like openPDS/SafeAnswers would help the data subject gain both knowledge and control of their inferred data being processed, and would solve the main problem with inferred data I described earlier: the complete lack of insight in when, what, how and by whom data are being inferred and processed.

5.3 Conclusion

At the beginning of my thesis I stated my main research question, see also section 1.1.1:

> **To what extent are the technical specifications of PDMs compatible with the relevant legal standards?**

which I then divided into three subquestions:

1. What EU Data Protection problem(s) can PDMs solve?
2. What EU Data Protection problem(s) can PDMs *not* solve?
3. What new problem(s) do PDMs create?

In this section, I take a final look at the technical features of PDMs and how they (can possibly) comply with the legal requirements discussed before.

5.3.1 Data minimization

The data minimization principle is one of the fundamental principles in each PDM I discussed here. Recalling section 3.2, the principle relates to Article 6.1(b) and (c) of the DPD and means that data controllers should only process personal data which are relevant and adequate for the specified purposes, as well as store data only for as long as necessary to fulfill those purposes.

We have seen how each PDM basically follows the same approach: the data subject selects which attributes they want to share and the PDM ensures that only those data are processed. The PDMs I discussed also ensure that - at least within the environment of the PDM - the relevant data are only shared with the data controller involved. IRMA for example implements the data minimization principle by letting the Scheme Manager (see 4.2.5) determine whether the necessary attributes for the specified purposes are proportionate, and then only lets the Relying Party verify those relevant attributes from the IRMA Card. Within the IRMA environment, storage is minimized to zero since Relying Parties do not store any attributes. This is an excellent example of data and storage minimization, although the latter remains tricky *outside* the environment of PDMs and cannot be guaranteed "once the data are out" of the PDM environment. See also section 5.3.2.

5.3.2 Storage minimization

We have seen it with every PDM discussed in this thesis: even when a PDM hypothetically both guarantees the security, integrity and confidentiality of the data and satisfies all legal requirements from the DPD and pGPDR, this guarantee only applies when *keeping the data inside the evironment of the PDM.* Once data controllers or other third parties store the data *outside* of the PDM (for example by screen capturing, saving decrypted data offline or writing data down), all guarantees are gone - and most likely the compliance with the legal requirements are too. Additional ("sticky") policies are required to regulate storage minimization.

5.3.3 Reuse of data

In their discussion about consent and control over personal data, [Whitley, 2011] noticed a change in the meaning of control:

> "Whilst in earlier times control over personal data may have been best undertaken by preventing the data from being disclosed, in an internet enabled society it is increasingly important to understand how disclosed data is being used and reused and what can be done to control this further use and reuse."

PDMs can do just that: give the data subject insight in what their data is being used for, and prevent further use and reuse of their data. Focusing on the latter (reuse of data), we have seen how - at least within the scope and environment of the PDMs discussed - a third party has to specify the purposes of data processing, after which the relevant, encrypted attributes of data are shared. As long as those attributes stay within the environment of the PDM, reuse is technically impossible thanks to sticky policies (Synergetics), one-time disclosure (IRMA), encryption (all) et cetera.

5.3.4 Dealing with observed and inferred data

We have also seen how some PDMs offer improvement when it comes to observed and inferred personal data: openPDS/SafeAnswers for example gives the user insight in what data are being inferred and (somehow) for what purposes. As we have seen, *insight* here basically means knowledge of the logic involved in inferring data, providing this knowledge to the user in an intelligible form and let the user erase and/or rectify inaccurate or inadequate data. By using inferring techniques themselves (for example using SafeAnswers modules), PDMs are potentially able to succeed in the former two: providing the data subject with an understandable translation of this logic (openPDS/SafeAnswers does not yet comply with all the requirements, but can rather be seen as a promising start in dealing with inferred data). The rectification and erasure of inaccurate data - related to the right to be forgotten as discussed in section 3.2 - seem harder to technically implement, as it requires some kind of accessible data storage which would have its own implications.

5.3.5 End-to-end-encryption

More about security than privacy related, yet very important: most PDMs implement end-to-end encryption. This ensures the integrity of the personal data "from upload to processing" and has two main advantages: the data are more valuable to third parties as they are guaranteed to be adequate and not tampered, and - since you need to obtain a decryption key to view the data - only concerned parties are able to actually view the contents of the (meanwhile decrypted) data.

5.3.6 Profiling vs. pseudonymity

As we have seen, PDMs can improve both quality (in terms of integrity, accuracy and such) and quantity (see also data minimization) in processing and

storing personal data. By pseudonymizing the data, PDMs make it harder for third parties to specifically target data subjects or use linkable data for predictive analyses: as mentioned in section 3.4, profiling on pseudonymous data is not presumed to be significantly affecting the interests, rights or freedoms of the data subject. However, as it becomes more and more obvious that pseudonymous data *can* in fact often be linked to a data subject, it also creates one of the biggest issues with PDMs: while the data become more valuable, PDMs cannot technically ensure that third parties do not combine or further process pseudonymous attributes.

Especially with observed and inferred data (see also the openPDS/SafeAnswers overview in section 4.4.5), PDMs have to be careful not to aggregate too many specific data about the data subject, which would then ultimately *increase* the possibility for third parties to profile their data subjects. This is definitely a problem PDMs create.

5.3.7 The human factor

Strictly speaking not a technical issue and thus without the scope of my research question, but still very important: human beings are the weakest link in every (technical) system. Even a technically perfectly designed PDM relies on the user to remember their own security code (for example a PIN in IRMA). With all personal data being stored in one place (Mydex, Synergetics), on one card (IRMA) or one phone (openPDS/SafeAnswers), the data subject suddenly has a large responsibility for the safety of their own personal data.

5.4 Discussion and future work

With this paper I have tried to present the current state of affairs regarding Personal Data Management systems. I have selected four PDMs for - what I believe to be - their features and relevance, dived into their technical framework as far as possible (or allowed) and compared their guarantees with the applicable legal framework which I explored earlier in the thesis. This led to both general and specific observations on what PDMs can mean for (digital) privacy and gaining back control on your own (online) profile.

In the previous section I have concluded the problems solved, not solved and created by PDMs. Note that I have not explicitly answered the three subquestions to my research question, as most issues are both solved and unsolved, and some solutions create new problems on the side. Nonetheless, I feel that I have answered the main research question by showing to what extent the technical possibilities of PDMs comply with the current legal framework.

Future research can be done with these same four PDMs, of which all four are still being developed, but also (or mainly) with new ones. Not all PDMs I discussed gave me full insight to their sources and systems: at the time of writing, Synergetics has yet to publish an article on their technologies which I expect to be a promising contribution to the world of PDMs. I do believe in the future of PDMs and am looking forward to see technical implementations that comply with the legal framework as much as possible, and ideally to actually see such a PDM in use on a large scale.

References

Alpár, G. and Hoepman, J.-H. (2013). A secure channel for attribute-based credentials. Radboud University Nijmegen, ICIS DS.

Alpár, G. and Jacobs, B. (2013). Credential design in attribute-based identity management. In *Bridging distances in technology and regulation, 3rd TILTing Perspectives Conference*, pages 189–204.

Brands, S. and Paquin, C. (2010). U-prove cryptographic specification v1. 0. *Microsoft Corporation*.

Bus, J. and Nguyen, M.-H. C. (2013). Personal data management a structured discussion. pages 270–288.

Camenisch, J. and Van Herreweghen, E. (2002). Design and implementation of the idemix anonymous credential system. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 21–30. ACM.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3.

de Montjoye, Y.-A., Shmueli1, E., Wang, S. S., and Pentland, A. S. (2014). openpds: Protecting the privacy of metadata through safeanswers.

Heath, W., Alexander, D., and Booth, P. (2013). Digital enlightenment, mydex, and restoring control over personal data to the individual. pages 253–269.

Hildebrandt, M., O'Hara, K., and Waidner, M. (2013). The value of personal data. digital enlightenment yearbook 2013.

Kellomäki, S. and Vervenne, L. (2013). end2end trust assurance voor stem van de patiënt.

Kuppinger, M. and Kearns, D. (2013). Life management platforms: Control and privacy for personal data. pages 243–252.

Nguyen, M.-H. C., Haynes, P., Maguire, S., and Friedberg, J. (2013). A user-centred approach to the data dilemma: Context, architecture, and policy. pages 227–242.

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57:1701.

Synergetics (2014). A contractual, accountability & governanceframework for an end2end trust assured patient-centric care management ecosystem.

Vullers, P. and Alpár, G. (2013). Efficient selective disclosure on smart cards using idemix. In *Policies and Research in Identity Management*, pages 53–67. Springer.

Whitley, E. A. (2011). Towards effective, consent based control of personal data. pages 165–179.

World Economic Forum (2012). Rethinking personal data: Strengthening trust. In *World Economic Forum*.

A Cited sources from the Article 29 Working Party

[1] Article 29 Data Protection Working Party, Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware, WP 17, 23.02.1999, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/1999/wp17_en.pdf

[2] Article 29 Data Protection Working Party, Opinion 15/2011 Consent, WP 187, 13.07.2011, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf

[3] Article 29 Data Protection Working Party, Working Document 01/2012 on epSOS, WP 189, 25.01.2012, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp189_en.pdf

[4] Article 29 Data Protection Working Party, Opinion 08/2012 providing further input on the data protection reform discussions, WP 199, 05.10.2012, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp199_en.pdf

[5] Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation, WP 203, 02.04.2013, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

[6] Article 29 Data Protection Working Party, Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation, 13.05.2013, http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130513_advice-paper-on-profiling_en.pdf

[7] Article 29 Data Protection Working Party, Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC", WP217, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

B Glossary

Data controller
The natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data. (source: Article 2 of the DPD)

Data minimization
The principle of "data minimization" means that a data controller should limit the collection of personal information to what is directly relevant and necessary to accomplish a specified purpose. They should also retain the data only for as long as is necessary to fulfil that purpose. In other words, data controllers should collect only the personal data they really need, and should keep it only for as long as they need it. (source: European Data Protection Supervisor[31])

Data processor
A natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller. (source: Article 2 of the DPD)

Data subject
One who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. (source: Article 2 of the DPD)

Inferred data
The result of "analysis of personal data" using data mining technologies on (collections of) volunteered and/or observed data. Examples are future consumption prediction techniques or credit scores. Inferred data need not necessarily be personal data, but might nonetheless have an impact on individuals. (source: [World Economic Forum, 2012])

Observed data
Data "captured by recording activities of users", such as cell phone location data, search histories or digital cookies. (source: [World Economic Forum, 2012])

Personal data
Any information relating to an identified or identifiable natural person. (source: Article 2 of the DPD)

Personal Data Management system
The user-centric management of an individual's own personal data facilitated by various types of architectures to make sure that a person can retain a degree of control over who gets access to which of her personal data. (source: [Hildebrandt et al., 2013])

[31] European Data Protection Supervisor - Glossary - `https://secure.edps.europa.eu/EDPSWEB/edps/site/mySite/pid/74`

Volunteered data

Data users "explicitly share about themselves". For example: pictures shared on social network sites (Facebook, Instagram, Twitter), personal blogs or billing information during an online purchase. (source: [World Economic Forum, 2012])

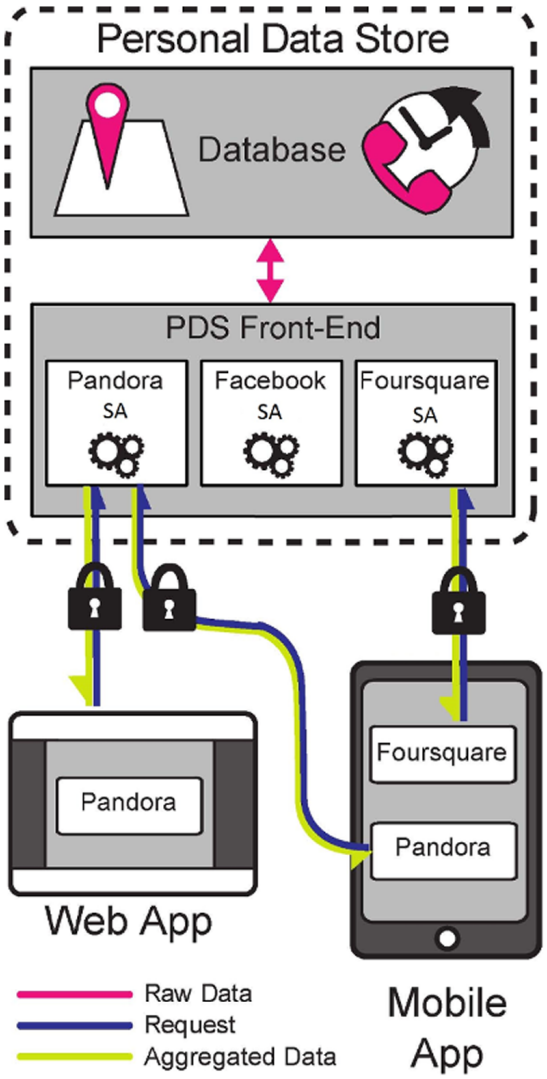

Figure 6: Architecture of openPDS/SafeAnswers with example SafeAnswers modules for Pandora, Facebook and Foursquare.

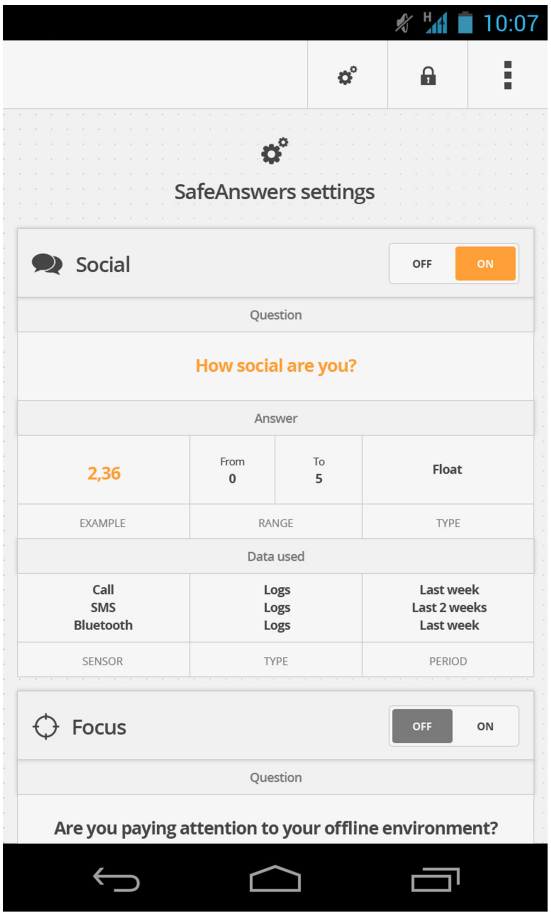

Figure 7: SafeAnswers example: this screen shows the question answered, examples of the possible responses, and the sensors used to compute the response.